cuk-H03666-10-P024842

ANNUAL REVIEW OF MEDICINE:

Selected Topics in
the Clinical Sciences

10

P. 24842

EDITORIAL COMMITTEE (1980)

C. H. COGGINS
Harvard Medical School

W. P. CREGER
Stanford University

R. M. DONALDSON, JR.
Yale University

E. W. HANCOCK
Stanford University

P. G. QUIE
University of Minnesota

R. M. ROSE
University of Texas

J. D. WILSON
University of Texas

Responsible for the organization of Volume 31
(Editorial Committee, 1978)

C. H. COGGINS
Harvard Medical School

W. P. CREGER
Stanford University

E. W. HANCOCK
Stanford University

P. G. QUIE
University of Minnesota

M. SIPERSTEIN
University of California Service
Veterans Administration Hospital,
San Francisco

J. D. CHERRY (Guest)
University of California

J. A. COLLINS (Guest)
Stanford University

K. TAYLOR (Guest)
Veterans Administration Hospital,
Palo Alto

Production Editor M. E. PLATT

ANNUAL REVIEW OF MEDICINE:

Selected Topics in the Clinical Sciences

WILLIAM P. CREGER, *Editor*
Stanford University School of Medicine

C. H. COGGINS, *Associate Editor*
Harvard Medical School

E. W. HANCOCK, *Associate Editor*
Stanford University School of Medicine

UNIVERSITY COLLEGE OF MEDICINE
LIBRARY
C. U.

VOLUME 31

1980

ANNUAL REVIEWS INC. 4139 EL CAMINO WAY PALO ALTO, CALIFORNIA 94306

ANNUAL REVIEWS INC.
Palo Alto, California, USA

COPYRIGHT © 1980 BY ANNUAL REVIEWS INC., PALO ALTO, CALIFORNIA, USA. ALL RIGHTS RESERVED. The appearance of the code at the bottom of the first page of an article in this serial indicates the copyright owner's consent that copies of the article may be made for personal or internal use, or for the personal or internal use of specific clients. This consent is given on the condition, however, that the copier pay the stated per-copy fee of $1.00 per article through the Copyright Clearance Center, Inc. (P. O. Box 765, Schenectady, NY 12301) for copying beyond that permitted by Sections 107 or 108 of the US Copyright Law. The per-copy fee of $1.00 per article also applies to the copying, under the stated conditions, of articles published in any Annual Review serial before January 1, 1978. Individual readers, and nonprofit libraries acting for them, are permitted to make a single copy of an article without charge for use in research or teaching. This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale.

REPRINTS The conspicuous number aligned in the margin with the title of each article in this volume is a key for use in ordering reprints. Available reprints are priced at the uniform rate of $1.00 each postpaid. The minimum acceptable reprint order is 5 reprints and/or $5.00 prepaid. A quantity discount is available.

International Standard Serial Number: 0066-4219
International Standard Book Number: 0-8243-0531-0
Library of Congress Catalog Card Number: A51-1659

2195
CI

Annual Reviews Inc. and the Editors of its publications assume no responsibility for the statements expressed by the contributors to this Review.

PRINTED AND BOUND IN THE UNITED STATES OF AMERICA

ANNUAL REVIEWS INC. is a nonprofit corporation established to promote the advancement of the sciences. Beginning in 1932 with the *Annual Review of Biochemistry,* the Company has pursued as its principal function the publication of high quality, reasonably priced Annual Review volumes. The volumes are organized by Editors and Editorial Committees who invite qualified authors to contribute critical articles reviewing significant developments within each major discipline.

Annual Reviews Inc. is administered by a Board of Directors whose members serve without compensation. The Board for 1980 is constituted as follows:

Dr. J. Murray Luck, Founder and Director Emeritus of Annual Reviews Inc.
Professor Emeritus of Chemistry, Stanford University
Dr. Joshua Lederberg, President of Annual Reviews Inc.
President, The Rockefeller University
Dr. James E. Howell, Vice President of Annual Reviews Inc.
Professor of Economics, Stanford University
Dr. William O. Baker, *President, Bell Telephone Laboratories*
Dr. Robert W. Berliner, *Dean, Yale University School of Medicine*
Dr. Sidney D. Drell, *Deputy Director, Stanford Linear Accelerator Center*
Dr. Eugene Garfield, *President, Institute for Scientific Information*
Dr. William D. McElroy, *Chancellor, University of California, San Diego*
Dr. William F. Miller, *President, SRI International*
Dr. Colin S. Pittendrigh, *Director, Hopkins Marine Station*
Dr. Esmond E. Snell, *Professor of Microbiology and Chemistry, University of Texas, Austin*
Dr. Harriet A. Zuckerman, *Professor of Sociology, Columbia University*

The management of Annual Reviews Inc. is constituted as follows:

John S. McNeil, Chief Executive Officer and Secretary-Treasurer
William Kaufmann, Editor-in-Chief
Charles S. Stewart, Business Manager
Sharon E. Hawkes, Production Manager
Ruth E. Severance, Promotion Manager

Annual Reviews are published in the following sciences: Anthropology, Astronomy and Astrophysics, Biochemistry, Biophysics and Bioengineering, Earth and Planetary Sciences, Ecology and Systematics, Energy, Entomology, Fluid Mechanics, Genetics, Materials Science, Medicine, Microbiology, Neuroscience, Nuclear and Particle Science, Pharmacology and Toxicology, Physical Chemistry, Physiology, Phytopathology, Plant Physiology, Psychology, Public Health, and Sociology. In addition, five special volumes have been published by Annual Reviews Inc.: *History of Entomology* (1973), *The Excitement and Fascination of Science* (1965), *The Excitement and Fascination of Science, Volume Two* (1978), *Annual Reviews Reprints: Cell Membranes, 1975–1977* (published 1978), and *Annual Reviews Reprints: Immunology, 1977–1979* (published 1980). For the convenience of readers, a detachable order form/envelope is bound into the back of this volume.

SOME RELATED ARTICLES IN OTHER *ANNUAL REVIEWS*

From the *Annual Review of Pharmacology and Toxicology,* Volume 19 (1979):

Salt in Hypertension and the Effects of Diuretics, Edward D. Freis

Nonsteroidal Anti-Inflammatory Agents, Rodney Nickander, F. Gilbert McMahon, and Anthony S. Ridolfo

Ethical Considerations in Cancer Chemotherapy, Emil J Freireich

From the *Annual Review of Physiology,* Volume 42 (1980):

Rheology and Hemodynamics, Giles R. Cokelet

From the *Annual Review of Public Health,* Volume 1 (1980):

Economic Evaluation of Public Health Programs, Lester B. Lave

Health and Disease in the United States, Lois A. Fingerhut, R. W. Wilson, and J. J. Feldman

Quality Assessment and Quality Assurance in Medical Care, Paul J. Sanazaro

Annual Review of Medicine
Volume 31, 1980

CONTENTS

[Note: Titles of chapters in Volumes 27–31 are arranged by category on pages 564–69.]

PREDICTION AND PREVENTION OF SUDDEN CARDIAC DEATH, *Arthur J. Moss* 1

SALT BALANCE AND LONG-TERM BLOOD PRESSURE CONTROL, *Arthur C. Guyton, Thomas G. Coleman, David B. Young, Thomas E. Lohmeier, and James W. DeClue* 15

HEMOGLOBIN A_{Ic} AND DIABETES MELLITUS, *Ronald J. Koenig and Anthony Cerami* 29

CONTROL OF ACID EXCRETION BY THE KIDNEY, *Richard L. Tannen* 35

AMANITA MUSHROOM POISONING, *D. H. Mitchel* 51

THE MEDICAL TREATMENT OF GALLSTONES, *Ian A. D. Bouchier* 59

INTESTINAL ABSORPTION AND MALABSORPTION OF FOLATES, *Charles H. Halsted* 79

PROSTAGLANDINS AND PLATELETS, *J. F. Mustard, R. L. Kinlough-Rathbone, and M. A. Packham* 89

HIGH DENSITY LIPOPROTEINS AND ATHEROSCLEROSIS, *G. J. Miller* 97

CLINICAL PROBLEMS OF VENTILATORY CONTROL, *Gernot B. Irsigler and John W. Severinghaus* 109

FLATULENCE, *Michael D. Levitt and John H. Bond* 127

MYOCARDIAL BIOPSY, *Chuichi Kawai, Akira Matsumori, and Keishiro Kawamura* 139

TREATMENT OF CHRONIC MYELOID LEUKEMIA, *P. A. Stryckmans* 159

PLASMA EXCHANGE IN GLOMERULONEPHRITIS AND RELATED VASCULITIDES, *C. M. Lockwood and D. K. Peters* 167

CLINICAL PHARMACOLOGY OF ACUTE ALLERGIC DISORDERS, *John C. Foreman and Lawrence M. Lichtenstein* 181

AUTOIMMUNE GRANULOCYTOPENIA, *G. L. Logue and D. S. Shimm* 191

NONINVASIVE EVALUATION OF THE CAROTID BRUIT, *Robert W. Barnes* 201

LEGIONNAIRES' DISEASE, *Richard D. Meyer and Sydney M. Finegold* 219

PULMONARY FUNCTION DURING AND AFTER COMMON RESPIRATORY INFECTIONS, *William J. Hall and R. Gordon Douglas, Jr.* 233

POSTURAL HYPOTENSION, *Michael G. Ziegler* 239

SNAKE VENOM POISONING IN THE UNITED STATES, *Findlay E. Russell* 247

EFFICACY AND SAFETY OF ORAL HYPOGLYCEMIC AGENTS, *Holbrooke S. Seltzer* 261

IDIOPATHIC MEMBRANOPROLIFERATIVE GLOMERULONEPHRITIS, *Youngki Kim and Alfred F. Michael* 273

PATHOLOGICAL CHANGES IN AORTOCORONARY BYPASS GRAFTS, *Grover M. Hutchins* 289

ERYTHROPOIESIS, *C. Peschle* 303

VASOPRESSIN FUNCTION IN THE SYNDROME OF INAPPROPRIATE ANTIDIURESIS, *R. Zerbe, L. Stropes, and G. Robertson* 315

EFFECTIVENESS AND RISKS OF CONTRACEPTION, *W. Droegemueller and R. Bressler* 329

PERICARDITIS AND RENAL FAILURE, *Roger Renfrew, Theodore J. Buselmeier, and Carl M. Kjellstrand* 345

THE ENDOTHELIUM AND THROMBOSIS, *Robert T. Wall and Laurence A. Harker* 361

GIARDIASIS, *James W. Smith and Martin S. Wolfe* 373

TREATMENT OF HYPERTHYROIDISM, *Michael R. McClung and Monte A. Greer* 385

BARTTER'S SYNDROME, *John R. Gill, Jr.* 405

PRIMARY PULMONARY HYPERTENSION, *Alfred P. Fishman and Giuseppe G. Pietra* 421

FOLATE-BINDING PROTEINS, *Neville Colman and Victor Herbert* 433

ACTIVE IMMUNIZATION FOR ADULTS, *Gilbert M. Schiff* 441

COMMUNITY CARDIOPULMONARY RESUSCITATION, *Leonard A. Cobb, Alfred P. Hallstrom, Robert G. Thompson, Lynn P. Mandel, and Michael K. Copass* 453

THE NATURAL HISTORY AND RESPONSE TO THERAPY OF LUPUS NEPHRITIS, *E. M. Ginzler, A. J. Bollet, and E. A. Friedman* 463

CONTROL OF CORONARY BLOOD FLOW, *Francis J. Klocke and Avery K. Ellis* 489

CONTROVERSIES IN PLATELET TRANSFUSION THERAPY, *Sherrill J. Slichter* 509

INFANT BOTULISM, *Stephen S. Arnon* 541

INDEXES

Cumulative Index of Contributing Authors, Volumes 27–31 561

Cumulative Index of Chapter Titles, Volumes 27–31 564

Ann. Rev. Med. 1980. 31:1–14
Copyright © 1980 by Annual Reviews Inc. All rights reserved

PREDICTION AND PREVENTION OF SUDDEN CARDIAC DEATH

❖7338

Arthur J. Moss, M.D.[1]

Heart Research Follow-up Program, Department of Preventive Medicine and Community Health and the Department of Medicine, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642

INTRODUCTION

An estimated one million cardiac deaths occur annually in the United States, and approximately one half of these are sudden. These numbers and percentages are imprecise because they are derived from a variety of epidemiologic and community studies using various definitions of sudden and often including non-cardiac causes of sudden death such as cerebral vascular disease, pulmonary embolism, and ruptured aortic aneurysm. Although sudden cardiac death (SCD) is common in such conditions as aortic valvular disease, cardiomyopathy, myocarditis, and the idiopathic long QT interval syndrome, and although SCD has been reported with mitral valve prolapse, preexcitation syndrome, and the various conduction defects, the most common cause of SCD is coronary heart disease.

During the past decade considerable progress has been made in the reduction of cardiac deaths from all causes, including SCD. A recent governmental vital statistics report has documented that the age-adjusted cardiovascular morality rates have declined about 32% over the past 30 years, with two-thirds of this decline occurring since 1968 (1). This reduced cardiac mortality rate is seen in both sexes, in whites and non-whites, and in every age group from the second to beyond the eighth decade. Experts in the field conclude that the recent decrease in both sudden and non-

---

[1] Supported in part by the Gebbie Foundation, Jamestown, New York, Merck Sharp & Dohme, West Point, Pennsylvania, and by grants HL-15790 and HL-22982 from the USPHS.

0066-4219/80/0401-0001$01.00

sudden mortality from coronary heart disease is real and not the result of artifacts or changes in death-certificate coding. Both primary prevention through changes in risk factors and better medical care have probably contributed to the decline (2).

These findings and conclusions suggest that further reduction in SCD will require a four-pronged approach: (*a*) primary prevention programs to reduce the incidence of cardiac disease in general and coronary heart disease in particular; (*b*) the development of more effective secondary prevention programs to reduce the percentage of patients dying with established heart disease; (*c*) the identification of groups and individuals who are at high risk of SCD; and (*d*) the selective application of medical treatment to high risk patients to reduce the likelihood of SCD.

DEFINITION OF SUDDEN

Various temporal yardsticks have been used for the definition of sudden death. In 1970, a broadly constituted international committee supported by the International Society of Cardiology and the American Heart Association agreed upon the following: "Sudden unexpected (natural) death is defined as death occurring instantaneously or within an estimated 24 hours of the onset of acute symptoms or signs" (3). This definition has been accepted by most epidemiologists, but clinicians feel uncomfortable with such a broad time range. Friedman et al (4) subdivided "sudden death" into instantaneous (prior to or within 30 seconds after the onset of any symptoms) and non-instantaneous (minutes to 24 hours after the onset of symptoms or signs) demise. Ruberman et al (5) in their follow-up studies of coronary patients in the Health Insurance Program of New York City defined sudden coronary death as that occurring within minutes of a patient's usual state of health in the absence of symptoms or findings suggesting acute myocardial infarction. Most investigators in the field utilize the one-hour definition. Myerburg (6) constructed a broad and useful definition of sudden cardiac death: "Natural death due to cardiac causes, occurring either instantaneously or up to one hour after the onset of symptoms in a patient who may or may not have known preexisting heart disease, but in whom the time and mode of death occur unexpectedly."

EPIDEMIOLOGIC STUDIES

Various epidemiologic studies have been carried out to determine the frequency and specific cause of sudden death in defined populations. Kuller et al (7) investigated the chronology and cause of death in a large sample (1857) of non-traumatic deaths in Baltimore residents between the ages of

20 and 65. Using a 24-hour definition for sudden, Kuller found that 32% of all non-traumatic deaths and 60% of all arteriosclerotic heart disease (ASHD) deaths were sudden. Furthermore, half of the ASHD sudden death victims had a history of heart disease prior to death and almost one quarter had seen a physician within a week prior to demise. In a subsequent postmortem study, Kuller found that SCD patients were characterized by a high frequency of prior heart disease, hypertension, diabetes, recent medical care, and prodromal symptoms (8). Also, SCD usually occurs at home or work and is rarely associated with an obvious acute precipitating event. Postmortem studies of these sudden ASHD deaths revealed heavy hearts (58% exceeded 450 g), a preponderance of severe and diffuse three- and four-vessel coronary atherosclerotic disease, and relatively infrequent acute lesions (fresh coronary thrombosis or recent myocardial infarction). Additional reports from the Baltimore group emphasized the combination of chronic myocardial damage and extensive multivessel disease as the substrate for SCD in ASHD; single-vessel coronary disease was infrequent and significant intramyocardial blood vessel disease was not demonstrated in this SCD group (9).

In the Michigan prospective epidemiologic investigation involving 8641 residents of Tecumseh (10), 46% (45 of 98) of the coronary heart disease deaths were sudden (within one hour of the onset of symptoms). In the majority of the sudden deaths, coronary heart disease or diabetes mellitus had been detected on prior examinations. Also, 84% of those dying suddenly had prior evidence of arrhythmias and conduction defects in a 12-lead electrocardiogram. The Tecumseh investigators concluded that although sudden coronary deaths occur without immediate prior warning, the victims are predisposed long before the catastrophic event.

In the so-called pooling project (the combined Albany and Framingham population studies), an investigation was performed to determine whether any characteristics differentiate coronary disease patients who die suddenly from those who die more slowly from the disease (11). The likelihood of sudden death could be estimated over a 14-fold range of risk from a multivariate profile of systolic blood pressure, relative body weight, cigarette habit, and electrocardiographic evidence of left ventricular enlargement. Although one third of the sudden deaths occurred in men in the topmost decile of risk, no combination of risk factors identified persons destined to die suddenly from those who would experience a more protracted death. That is, the risk profile is essentially the same whether the terminal event is sudden and unexpected or prolonged.

Social and behavioral factors have been implicated in sudden coronary deaths. Although many anecdotal experiences have been reported (12–14), few scientific studies substantiate such a relationship. A recent investigation of sudden death during the posthospital phase of myocardial infarction in a large prepaid group practice (Health Insurance Plan of Greater New York) substantiated a significant relation between low education level and SCD (15). During a three-year follow-up period after myocardial infarction, men with eight years of schooling or less who had complex ventricular premature beats in a one-hour monitoring period had over three times the risk of sudden coronary death found among better-educated men with the same arrhythmia. This relationship was independent of traditional risk factors and the usual clinical characteristics affecting prognosis.

PATHOLOGIC STUDIES

Over 20 years ago Spain et al (16) demonstrated that 80% of patients with sudden deaths had significant coronary stenosis in more than one vessel. In only 15% of Spain's cases was the main lesion limited to one vessel, usually the left main or left anterior descending arteries. (The remaining 5% of patients had nonsignificant coronary stenosis.) Extensive coronary atherosclerotic disease has been substantiated as the substrate for SCD in numerous subsequent reports (17–19). Despite the diffuse nature of the coronary atherosclerotic process in SCD, only one third of such patients have occlusive coronary artery thrombi (20). In contrast, patients dying non-suddenly with transmural myocardial infarction have a high frequency (90%) of occlusive thrombi. When organizing thrombi are observed in SCD, the sites at which they occur show a high prevelance of acute lesions—plaque rupture and intramural hemorrhage (20). Friedman et al (4) described different pathologic findings in patients dying instantaneously (<30 seconds after onset of any symptoms) from those dying suddenly (within minutes to 24 hours). Patients dying instantaneously had extensive chronic coronary arterial disease but almost no evidence of an acute pathologic lesion. In contrast, significant acute coronary lesions (thrombosis, plaque rupture, intramural hemorrhage) were found in 82% of the non-instantaneous sudden coronary deaths. These findings suggest that instantaneous coronary death is an electrical event, possibly caused by acute myocardial ischemia, but not the result of acute coronary thrombosis.

Some investigators speculate that the lethal cardiac dysrhythmias in SCD may arise from ischemic necrosis involving the specialized Purkinje conducting system. Lie & Titus (21) carried out a detailed histopathologic study of the conducting system in 49 SCD victims (within six hours of the onset of acute symptoms). Acute ischemic necrosis of the Purkinje fibers was an exceptional finding, and there were no specific histopathologic lesions of the conduction system peculiar to SCD. A seemingly intact conduction system was not inconsistant with the development of lethal cardiac arrhythmias. Lie & Titus concluded that SCD probably resulted from electrical instability of the acutely ischemic contractile myocardium rather than a direct injury to the specialized conductive tissue of the heart.

A recent study from Helsinki focused its attention on the presence of recent myocardial infarction and the prevalence of different coronary lesions or their combination in relation to the time course of the fatal attack (19). Among 151 SCD victims (<24 hours), necropsy revealed a recent myocardial infarction in 77% (51% with "early" changes and 26% with "definite" findings using the WHO classification scheme). The clinical suddenness of death did not relate to the stage of the recent myocardial lesion (early vs definite). In about half of the patients with an early or definite infarction, as well as those with no recent infarction at necropsy, the fatal attack ended abruptly (<2 hours). However, patients with non-abrupt SCD (2–24 hours) were more likely to have definite rather than early infarction. Also, the severity of the obstructive changes in the coronary arteries was no different in patients who died instantaneously than in those who died after a period of symptoms, whether under or over two hours. Thus, suddenness of cardiac death is not associated with a specific pathologic pattern in either the myocardium or the coronary arteries.

RESUSCITATED SCD VICTIMS

Many communities have established emergency mobile medical care systems, and during the past decade many patients have been resuscitated from out-of-hospital cardiac arrest. This resuscitated group presents a unique opportunity to study the pathophysiologic factors responsible for lethal arrhythmias. Weaver et al (22) reported the angiographic findings in 64 patients resuscitated from out-of-hospital ventricular fibrillation (VF). The majority of VF patients (72%) had a previous history of cardiovascular disease, while in 28% ventricular fibrillation was the first manifestation of coronary artery disease. Advanced coronary atherosclerosis was a frequent finding: 94% of the patients had severe stenosis in one or more of the major coronary arteries, and 70% had ventricular contraction abnormalities. Of the 64 resuscitated patients, 14 developed a second potentially lethal arrhythmic episode during an average follow-up of 20 months. Patients who "died the second time" had more severe triple-vessel coronary disease, lower ejection fractions, and much more severe ventricular wall motion abnormalities than did survivors. These findings highlight the importance of combined coronary and myocardial factors in the initiation and recurrence of ventricular fibrillation.

Other studies from the Seattle group provide insight into the mechanisms responsible for the terminal event. In 1974 the Seattle group reported on the posthospital follow-up of 146 patients resuscitated from prehospital VF (23). Acute transmural myocardial infarction was confirmed in only 17% of the patients, and myocardial necrosis (LDH-isoenzyme criteria) was observed in 33%; the remaining 50% of the resuscitated victims showed no evidence of acute myocardial infarction. This latter non-infarctive group had a two-year mortality more than three times that of the patients with acute infarction following the initial resuscitation. Thus, the phenomenon of SCD should not be equated with acute myocardial infarction. Also, patients resuscitated from ventricular fibrillation without associated acute myocardial infarction are prone to sudden death—most probably from residual myocardial fibrosis and ischemia, which contributes to continuing electrical instability. Those with acute myocardial infarction, on the other hand, apparently have only a temporary electrical instability.

POSTHOSPITAL PHASE OF MYOCARDIAL INFARCTION

Patients who survive the acute phase of myocardial infarction have a one-year posthospital mortality of approximately 10%, with a continued mortality rate of about 3 to 5% per year thereafter. Posthospital cardiac death may be sudden or non-sudden with a heterogeneous combination of arrhythmic, mechanical, and ischemic factors playing a role in the terminal event. In 1972, a prospective postcoronary follow-up program was initiated in Rochester, New York, to detail the posthospital course of patients surviving an acute myocardial infarction. From 1973 to 1976, 978 patients less than 66 years of age were enrolled in the study. Prior to hospital discharge, a detailed clinical evaluation was carried out including a 6-hour Holter ECG tape recording. The average age of these patients was 53.7 years; 81% were male and 40% had an anterior myocardial infarction. During an average follow-up of almost three years, 117 patients died of cardiac cause. Among the witnessed deaths ($n = 98$), 56% were sudden (≤ 1 hour).

The Rochester postinfarction population was analyzed to evaluate the chronology (interval from hospital discharge to demise) and suddenness (elapsed time from acute cardiac symptoms to demise) of cardiac death. During the first postinfarction year, 57% of the non-sudden cardiac deaths occurred within two months after hospital discharge, and 77% of the SCD (≤ 1 hour) occurred within 3 to 12 months ($P < 0.02$). No predischarge clinical variables, including Holter-recorded ventricular premature beats, distinguished between subsequent sudden and non-sudden cardiac death. These findings indicate that the time course of the terminal event is more difficult to predict than had been appreciated previously.

Numerous investigators find that certain ventricular premature beat (VPB) characteristics recorded in patients following acute myocardial infarction are harbingers of subsequent cardiac death (24–26). Recently, investigators from the Health Insurance Program (HIP) in New York City demonstrated that men with complex VPBs (R on T, runs of two or more, multiform, or bigeminal beats) experienced a risk of sudden death three times that of men with simple or no VPBs (5). Furthermore, complex VPBs made an independent contribution to the mortality risk. We recently carried out a similar analysis on the Rochester population using the same criteria for complex VPBs as the HIP study. Cardiac mortality was categorized as sudden ($\leq$1 hour) or non-sudden ($>$1 hour). Complex VPBs were associated with a significantly increased cardiac death rate, but the complex patterns did not discriminate between sudden and non-sudden death (27). These findings indicate that Holter-recorded complex VPBs lack specificity with regard to the prediction of the mode of cardiac death.

Various studies in progress or recently reported provide additional clues to who is at risk in the postmyocardial infarction population. In the Stanford Rehabilitation Program (28), treadmill exercise testing was performed three weeks after the acute event in uncomplicated myocardial infarction patients at risk for a subsequent cardiac event. Prognosis was more closely related to exercise-induced ischemic abnormalities (ST segment depression $\geq$0.2 mV) and to mechanical dysfunction (low maximal work load) than to electrical instability (VPBs). In a more representative postinfarction population, Weld et al (29) substantiated the importance of poor mechanical performance as a marker of subsequent cardiac death. Schulze et al (30) performed Holter monitoring and gated cardiac blood pool scanning on patients prior to hospital discharge following myocardial infarction. All sudden deaths occurring within a mean follow-up time of seven months developed in patients with the combination of a depressed ejection fraction ($<$0.4) and complex VPBs. Recently, Greene et al (31) utilized the repetitive ventricular response (two or more VPBs produced by a single ventricular pacing stimulus) to evaluate posthospital outcome in 48 survivors of myocardial infarction. Of those who had a repetitive response to the pacing stimulus, 79% (15 of 19) had symptomatic ventricular tachycardia or SCD within a 12-month follow-up. Schwartz & Wolf (32) showed in a group of 55 patients with myocardial infarction that the risk for SCD was 2.2 times greater with a prolonged QT_c ($>$ 0.44 sec) than with a normal QT_c. Finally, two-dimensional echocardiography is now being used to identify high risk postinfarction patients. Eaton et al (33) used two-dimensional echo to show that coronary patients with regional infarct expansion have a significantly greater eight-week mortality than patients without this finding.

CIRCUMSTANCES ATTENDING SCD

Although a large amount of data has been accumulated regarding the epidemiology, pathology, and risk factors pertaining to SCD, there is little information describing the circumstances of the terminal event. Kuller, in his epidemiologic study of SCD in Baltimore, noted that the onset of the terminal event was usually at home or work and rarely associated with an obvious acute precipitating event (7). Various authors have speculated on the potentially important role of emotional and psychologic factors in initiating SCD, but appropriate studies to support or refute these speculations have not been reported. Lown (34) has investigated in the animal model transient neurologic and psychologic factors that can lower the threshold for ventricular fibrillation in the vulnerable heart. In fact, Lown believes that investigative focus into the factors responsible for SCD should be shifted from the heart as target to the brain as trigger.

In light of the aforementioned comments, the report by Myers & Dewar (14) into the circumstances attending 100 male sudden deaths is of interest. These investigators paid particular attention to the patient's normal habits and mode of life as well as the circumstances in which the patient died to determine whether these circumstances were unusual or of more than coincidental significance. Each subject served as his own control in terms of long uneventful periods that preceded the terminal event; in addition, 100 non-fatal infarct patients served as an accessory control group. Very strenuous exercise, the season of the year, the environmental temperature or its recent change, and chronic psychologic stress were similarly distributed in the SCD and non-fatal infarct controls. Moderate physical activity, evening hours (6–10 PM), Saturday, a recent meal accompanied by alcohol, and acute psychologic stress were significantly more frequent in the SCD group than in the non-fatal controls. These findings are compatible with the thesis that activation of the sympathetic nervous system and the neurogenic release of catecholamines are important factors in the genesis of SCD.

PREDICTORS OF SCD

Prediction of SCD is difficult, and the available risk factors lack specificity and sensitivity in victim identification and the chronologic occurrence of the event. Classifying patients by the type of clinical cardiac disease provides some insight into the predictors of SCD.

Coronary Heart Disease

Within an asymptomatic population, 22% of new coronary events are SCD. Available evidence indicates that the traditional risk factors for coronary disease (hypertension, hypercholesterolemia, and cigarette smoking) are sensitive but non-specific predictors of SCD. In patients with established coronary heart disease (angina, prior myocardial infarction, and positive exercise test), the severity of the mechanical, electrical, ischemic, and coronary abnormalities bears a direct relationship to the risk of both sudden and non-sudden cardiac death. Recent invasive studies using ventricular pacing stimuli to provoke a repetitive ventricular response appear to hold some promise in identifying postcoronary patients at high risk of SCD (31).

Valvular Aortic Stenosis

Patients with rheumatic or congenital aortic stenosis, symptoms of angina or congestive heart failure, and a critically reduced aortic valve area (< 1.0 cm^2) experience a high mortality within a two-year interval after onset of cardiac deterioration. Syncope is particularly ominous and usually reflects transient hemodynamic disturbances with secondary premalignant arrhythmias.

Hypertrophic Cardiomyopathy

Ventricular arrhythmias are common in patients with hypertrophic cardiomyopathy of the obstructive (IHSS, HOCM)[2] and non-obstructive types, and VPBs may play a causal role in the higher incidence of SCD observed in these conditions. The combination of a family history of SCD, ventricular hypertrophy by ECG, young age, and frequent VPBs identifies the individual at excess risk for SCD.

Mitral Valve Prolapse

Mitral valve prolapse with click-murmur findings and frequent VPBs is a common clinical condition. Recent longitudinal studies substantiate the benign nature of this condition in most patients. However, a small subset of patients with a history of syncope or presyncope, blood relatives with this syndrome, ECG ST-T wave changes including slight QT prolongation, and frequent or complex VPBs on Holter recordings is vulnerable to SCD from ventricular fibrillation. This subgroup warrants close observation and aggressive antiarrhythmic therapy.

Long QT Syndrome

Prolongation of the QT interval is associated with ventricular arrhythmias, syncope, and sudden death (35). The delayed repolarization phenomenon has been categorized into primary or idiopathic types with and without heritable features and acquired forms secondary to drugs, electrolyte imbalance, and cerebral vascular disease. In many patients with long QT syndrome (LQTS), VPBs are present on 24-hour recordings, and transient ventricular flutter or fibrillation has been recorded during syncopal episodes. The familial LQTS with VPBs is associated with a high SCD rate, estimated at 73% over an average five-year period.

---

[2] IHSS = idiopathic hypertrophic subaortic stenosis; HOCM = hypertrophic obstructive cardiomyopathy.

Atrioventricular Conduction Disturbance

Atrioventricular conduction disturbance (AVCD) with varying degrees of heart block may develop from a variety of different disease conditions (degenerative, arteriosclerotic, inflammatory, metabolic, etc). Although PR interval prolongation and bundle and fascicular conduction disturbances are common, His bundle recordings have not been useful in identifying patients at risk of asystolic SCD. Syncope of the classical Morgagni-Adams-Stokes type is the hallmark of impending SCD. Pacemaker therapy is indicated in patients with syncope complicating AVCD.

PREVENTION OF SCD

As pointed out in the previous sections, SCD may occur with a variety of cardiovascular disorders, yet coronary heart disease accounts for over 90% of all SCD events. Any attempt to prevent SCD must focus on the problems of coronary heart disease. Prevention is an epidemiologic term frequently applied to the eradication of communicable and contagious disease, and it may be inappropriate to think in similar terms of prevention of SCD due to coronary atherosclerosis. Rather, emphasis should be placed on reducing the occurrence of SCD through a variety of therapeutic and prophylactic measures.

Community Measures

Many communities have substantiated the life-saving effectiveness of rapidly responsive mobile units capable of on-the-spot defibrillation. The resuscitation rate of out-of-hospital cardiac arrest has been enhanced when lay members of the society have provided temporizing cardiopulmonary resuscitation (CPR) with external cardiac massage and mouth-to-mouth resuscitation. Thus, local American Red Cross and American Heart Association chapters should initiate community-wide training of lay personnel in CPR and should encourage the development of mobile resuscitation units.

Public Health Measures

SCD is frequently the initial manifestation of coronary heart disease, and extensive coronary atherosclerotic disease is usually present on autopsy examination. It follows that therapeutic measures that prevent, inhibit, or reverse the atherosclerotic process provide the greatest long-term likelihood of reducing the rate of SCD. In this regard, the progressive reduction in atherosclerotic cardiac mortality over the past ten years suggests that continuing efforts should be directed toward the control of the primary risk factors. Reduction in tobacco usage, blood pressure, and blood lipids together with augmented physical activity should reduce the risk of coronary atherosclerotic disease in the general population with a resultant diminution in the occurrence of SCD.

Physician-Initiated Measures

Experience in Coronary Care Units (CCU) substantiates the benefits of aggressive therapy in reducing mortality from acute myocardial infarction. The fundamental principles involved in the CCU approach are (*a*) risk stratification (admission of high risk patients with acute myocardial ischemia or infarction to specialized care units) and (*b*) the application of selective prophylactic and therapeutic measures (antiarrhythmic agents, defibrillation, hemodynamic stabilization, etc). Similar principles should be used by the physician when attempting to prevent out-of-hospital SCD.

RISK STRATIFICATION High risk groups with established coronary heart disease include postinfarction patients during the first year after hospital discharge, patients with persistent angina pectoris (particularly those with an unstable pattern), and individuals resuscitated from out-of-hospital cardiac arrest. Further subsets of risk can be identified with activity evaluation testing (those with reduced functional performance on treadmill), Holter monitoring (complex VPBs), and radionuclide cineangiography (decreased ejection fraction, dyskinetic wall motion). Low risk groups free of complex ventricular rhythms and with good performance and cardiac function can be eliminated from prophylactic therapies since their mortality rate is remarkably low over extended periods of time (36).

PROPHYLACTIC AND THERAPEUTIC MEASURES Specific interventions should be initiated according to the risk mechanisms most likely to be operative in the individual patient. Multiple risk factor interventions are usually required because of the complex interplay of diverse pathophysiologic mechanisms that result in SCD in patients with extensive coronary disease.

Pharmacologic interventions Although postinfarction trials with beta blockers (practolol, alprenolol) have shown promising results in reducing posthospital mortality (37, 38), a recent propranolol trial in England failed to confirm the earlier studies (39). Current physiologic recommendations are (*a*) propranolol therapy for patients with manifest myocardial ischemia (angina pectoris, ST and T wave abnormalities with exercise testing); (*b*) afterload-reducing drugs for patients with significant left ventricular dysfunction (pulmonary congestion, poor activity performance, significantly reduced ejection fraction); and (*c*) full-dose therapy with antiarrhythmic agents in patients with complex, life-threatening, malignant-type ventricular rhythms. These pharmacologic interventions should be applied for limited periods during the intervals of enhanced risk (usually the first 6 to 12 months after the coronary event).

Surgical interventions Coronary artery bypass surgery has produced a significant reduction in cardiac mortality in symptomatic patients with left main coronary stenosis and possibly in those with trivessel coronary disease. Various surgical therapies including interruption of the myocardial reentry circuit, sympathectomy, stellate ganglionectomy, and pacemakers have proven efficacious for rhythm stabilization in selected patients with recurrent malignant ventricular arrhythmias unresponsive to conventional drug therapy. An innovative approach utilizing a newly developed implanted demand ventricular defibrillator is currently being tested (40). Such a device would have particular applicability in high risk patients with recurrent ventricular defibrillation.

SUMMARY

The most common cause of SCD is coronary heart disease. Epidemiologic and pathologic studies have demonstrated the presence of chronic myocardial damage and extensive multivessel coronary disease as the substrate for SCD in ASHD. Instantaneous coronary death is an electrical event and not the result of acute coronary thrombosis. Patients resuscitated from out-of-hospital cardiac arrest who do not evolve acute myocardial infarction are at a greater risk of subsequent SCD than are infarctive patients. In studies of the posthospital phase of myocardial infarction, there are no predictive clinical variables that distinguish between sudden and non-sudden cardiac death. Current investigations suggest that activation of the sympathetic nervous system and the neurogenic release of catecholamines are important triggering factors in the genesis of arrhythmic SCD. Reduction in the occurrence of SCD will require community (mobile resuscitation units), public health (control of primary risk factors), and physician-initiated (antiarrhythmic therapy) measures.

Literature Cited

1. United States Department of Health, Education, and Welfare. 1978. Monthly Vital Statistics Report: Annual summary for the United States 1977. *DHEW Publ. No. [PHS] 79–1120.* Washington, DC: GPO
2. Levy, R. I. 1979. Stroke decline: implications and prospects. *N. Engl. J. Med.* 300:490–91
3. Paul, O., Schatz, M. 1971. On sudden death. *Circulation* 43:7–10
4. Friedman, M., Manwaring, J. H., Rosenman, R. H., Donlon, G., Ortega, P., Grube, S. M. 1973. Instantaneous and sudden deaths: clinical and pathological differentiation in coronary artery disease. *J. Am. Med. Assoc.* 225:1319–28
5. Ruberman, W., Weinblatt, E., Goldberg, J. D., Frank, C. W., Shapiro, S. 1977. Ventricular premature beats and mortality after myocardial infarction. *N. Engl. J. Med.* 297:750–57
6. Myerburg, R. J. 1978. Sudden death. *J. Contin. Educ. Cardiol.* 14:15–29
7. Kuller, L., Lilienfeld, A., Fisher, R. 1966. Epidemiological study of sudden and unexpected deaths due to arteriosclerotic heart disease. *Circulation* 34: 1056–68
8. Kuller, L., Cooper, M., Perper, J. 1972. Epidemiology and sudden death. *Arch. Intern. Med.* 129:714–19
9. Perper, J. A., Kuller, L. H., Cooper, M. 1975. Arteriosclerosis of coronary arteries in sudden unexpected deaths. *Circulation* 51,52(III):27–33
10. Chiang, B. N., Perlman, L. V., Fulton, M., Ostrander, L. D., Epstein, F. H. 1970. Predisposing factors in sudden cardiac death in Tecumseh, Michigan: a prospective study. *Circulation* 41:31–37
11. Doyle, J. T., Kannel, W. B., McNamara, P. M., Quickenton, P., Gordon, T. 1976. Factors related to suddenness of death from coronary disease: combined Albany-Framingham studies. *Am. J. Cardiol.* 37:1073–78
12. Engel, G. L. 1971. Sudden and rapid death during psychological stress: folklore or folkwisdom? *Ann. Intern. Med.* 74:771–82
13. Engel, G. L. 1978. Psychologic stress, vasodepressor (vasovagal) syncope, and sudden death. *Ann. Intern. Med.* 89: 403–12
14. Myers, A., Dewar, H. A. 1975. Circumstances attending 100 sudden deaths from coronary artery disease with coroner's necropsies. *Br. Heart J.* 37: 1133–43
15. Weinblatt, E., Ruberman, W., Goldberg, J. D., Frank, C. W., Shapiro, S., Chaudhary, B. S. 1978. Relation of education to sudden death after myocardial infarction. *N. Engl. J. Med.* 299:60–65
16. Spain, D. M., Bradess, V. A. 1960. The relationship of coronary thrombosis to coronary atherosclerosis and ischemic heart disease. *Am. J. Med. Sci.* 240: 701–10
17. Schwartz, C. J., Gerrity, R. G. 1975. Anatomical pathology of sudden unexpected cardiac death. *Circulation* 51,52(III):18–26
18. Bashe, W. J., Nobuhisa, B., Keller, M. D., Geer, J. C., Anthony, J. R. 1975. Pathology of atherosclerotic heart disease in sudden death: the significance of myocardial infarction. *Circulation* 51,52(III):63–69
19. Rissanen, V., Romo, M., Siltanen, P. 1978. Prehospital sudden death from ischaemic heart disease: a postmorten study. *Br. Heart J.* 40:1025–33
20. Baba, N., Bashe, W. J., Keller, M. D., Geer, J. C., Anthony, J. R. 1975. Pathology of atherosclerotic heart disease in sudden death: organizing thrombosis and acute coronary vessel lesions. *Circulation* 51,52(III):53–59
21. Lie, J. T., Titus, J. L. 1975. Pathology of the myocardium and the conduction system in sudden coronary death. *Circulation* 51,52(III):41–52
22. Weaver, W. D., Lorch, G. S., Alvarez, H. A., Cobb, L. A. 1976. Angiographic findings and prognostic indicators in patients resuscitated from sudden cardiac death. *Circulation* 54:895–900
23. Baum, R. S., Alvarez H. III, Cobb, L. A. 1974. Survival after resuscitation from out-of-hospital ventricular fibrillation. *Circulation* 50:1231–35
24. The Coronary Drug Project Research Group. 1973. Prognostic importance of premature beats following myocardial infarction: experience in the coronary drug project. *J. Am. Med. Assoc.* 223: 1116–24
25. Kotler, M. N., Tabatznik, B., Mower, M. M., Tominaga, S. 1973. Prognostic significance of ventricular ectopic beats with respect to sudden death in the late postinfarction period. *Circulation* 47: 959–66
26. Vismara, L. A., Amsterdam, E. A., Mason, D. T. 1975. Relation of ventricular arrhythmias in the late hospital phase of acute myocardial infarction to sudden death after hospital discharge. *Am. J. Med.* 59:6–12

27. Moss, A. J., Davis, H. T., DeCamilla, J., Bayer, L. W. 1979. Ventricular ectopic beats and their relation to sudden and non-sudden cardiac death after myocardial infarction. *Circulation.* In press
28. DeBusk, R. F. 1979. Personal communication. Stanford Univ. Sch. Med., Stanford, Calif. 94305
29. Weld, F. M., Bigger, J. T. 1978. Risk stratification by exercise testing and 24-hour ambulatory electrocardiogram 2 weeks after acute myocardial infarction. *Circulation* 57,58(II):198
30. Schulz, R. A., Strauss, H. W., Pitt, B. 1977. Sudden death in the year following myocardial infarction: relation to ventricular premature contractions in the late hospital phase and left ventricular ejection fraction. *Am. J. Med.* 62: 192–99
31. Greene, H. L., Reid, P. R., Schaeffer, A. H. 1978. The repetitive ventricular response in man: a predictor of sudden death. *N. Engl. J. Med.* 299:729–34
32. Schwartz, P. J., Wolf, S. 1978. QT interval prolongation as predictor of sudden death in patients with myocardial infarction. *Circulation* 57:1074–77
33. Eaton, L. W., Weiss, J. L., Bulkley, B. H., Garrison, J. B., Weisfeldt, M. L. 1979. Regional cardiac dilatation after acute myocardial infarction: recognition by two-dimensional echocardiography. *N. Engl. J. Med.* 300:57–62
34. Lown, B. 1979. Sudden cardiac death: the major challenge confronting contemporary cardiology. *Am. J. Cardiol.* 43:313–28
35. Moss, A. J., Schwartz, P. J. 1979. Sudden death and the idiopathic long Q-T syndrome. *Am. J. Med.* 66:6–7
36. Moss, A. J., Davis, H., DeCamilla, J. 1978. Survivorship patterns after myocardial infarction: implications for intervention trials. *Circulation* 58(II):152
37. Wilhelmsson, C., Vedin, J., Wilhelmsen, L., Tibblin, G., Werko, L. 1974. Reduction of sudden deaths after myocardial infarction by treatment with alprenolol. *Lancet* 2:1157–62
38. Multicenter International Study. 1975. Improvement in prognosis of myocardial infarction by long-term beta-adrenoreceptor blockade using practolol. *Br. Med. J.* 3:735–39
39. Baber, N. 1979. Personal communication. Imperial Chem. Ind., Mereside Alderley Park, Macclesfield Cheshire, SK10 4 TG, England
40. Mirowski, M. 1979. Personal communication. Sinai Hosp. Baltimore, Md. 21215

Ann. Rev. Med. 1980. 31:15–27
Copyright © 1980 by Annual Reviews Inc. All rights reserved

SALT BALANCE AND LONG-TERM BLOOD PRESSURE CONTROL

❖7339

Arthur C. Guyton, M.D., Thomas G. Coleman, M.D.,
David B. Young, M.D., Thomas E. Lohmeier, M.D.,
and James W. DeClue, M.D.

Department of Physiology and Biophysics, University of Mississippi
Medical School, Jackson, Mississippi 39216

Most of us are well aware of the many studies showing that segregated groups of people who live with meager intakes of salt usually maintain entirely normal blood pressures throughout their lives (1–4). At the other extreme, populations in which salt intake is greater than the norm have an especially high incidence of hypertension (5, 6). One of the most graphic examples of relationships between salt intake and blood pressure was recorded by Prior and his colleagues (1, 2): the natives living on the island of Pukapuka and some other Pacific islands have a daily salt intake many times less than that of persons living in the Western culture, and their arterial pressures do not increase as they age. Yet, many of these natives have migrated to New Zealand and have taken up the cultural and dietary habits of the European New Zealanders. Subsequently, these migrants experience an incidence of increasing arterial pressure similar to that in the other new Zealanders of different racial stock.

SHORT-TERM SALT LOADING AND HYPERTENSION

It is common experience that a person can eat tremendous amounts of salt for a few days or even a few weeks at a time without experiencing a significant rise in arterial pressure; or he can go for weeks with almost no

0066-4219/80/0401-0015$01.00

salt intake, also without a significant change in pressure. These effects have been quantitated in the past few years by Langston et al (7), Douglas et al (8), Coleman & Guyton (9), DeClue et al (10), and Manning et al (11). The curve labeled "CHRONIC" in Figure 1 illustrates an average drawn from data in these studies, and shows the relationship of arterial pressure to urinary sodium output when the sodium intake is increased progressively over a period of two weeks from about 0.2 of normal up to 8 times normal. Note that the arterial pressure hardly changes, though the sodium output from the kidneys increases markedly. This curve is called a *long-term renal function curve*: It illustrates that, whatever the intake load of sodium within very wide limits, the kidney can excrete the salt load with very slight change in arterial pressure. Similar quantitative studies in human beings (12) show that the arterial pressure in the normal person changes an average of only

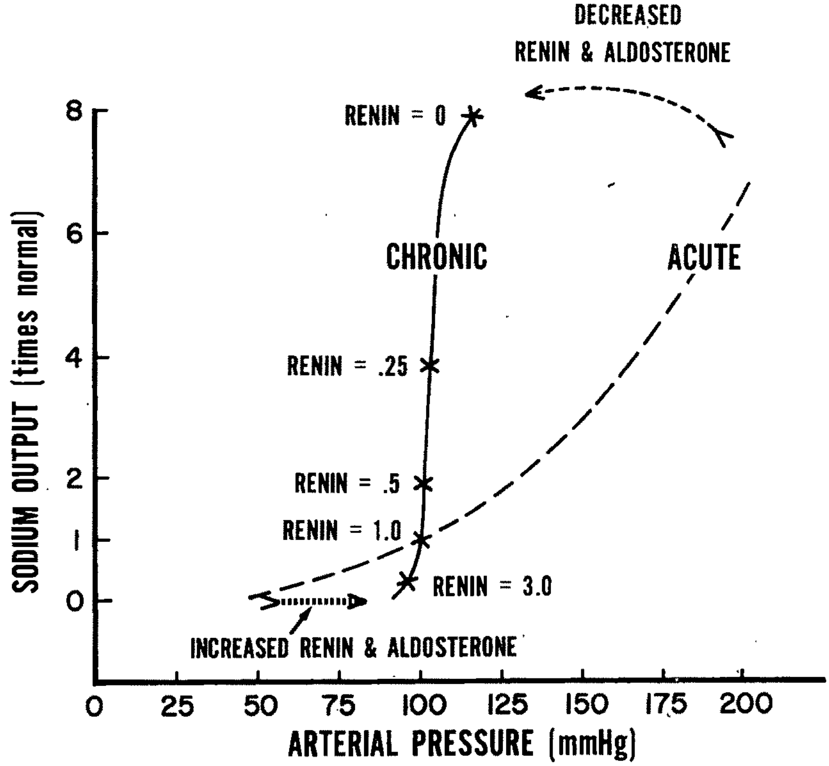


Figure 1 The normal acute and chronic renal function curves, illustrating the relationship between arterial pressure and urinary sodium output. The dashed arrows illustrate the directional shift of the acute curve toward the chronic curve caused by renin and aldosterone. The renin values are expressed in relative units with 1.0 equal to the normal value.

17 mm Hg when the salt intake is varied between 0.1 of normal salt intake and 10 times normal intake, a range of 100-fold.

This relationship between arterial pressure and urinary sodium output is very important in understanding the relationship between salt load and blood pressure. Thus we need to examine the physiological causes for the extremely steep relationship between pressure and sodium output as illustrated by the chronic renal function curve of Figure 1. The curve in Figure 1 that is labeled "ACUTE" will help explain this. This curve shows the relationship between arterial pressure and sodium output found in acute laboratory experiments in which the kidney is perfused with blood and the arterial pressure is changed through a range of values (13–15). One can readily see the marked difference between the two curves. But the conditions of the two experiments also are markedly different. For the acute curve, the arterial pressure was increased over a period of a few minutes to an hour or more. For the chronic curve, the salt intake was increased progressively over a period of several weeks; this caused a slight rise in arterial pressure but it also caused other physiological changes. For instance, it is well known that increasing the sodium intake decreases both renin and aldosterone secretion, while decreasing the sodium intake has the opposite effects. DeClue and his associates (10) and Young and his associates (16, 17) demonstrated the effects illustrated by the two dashed arrows in Figure 1: the decreased renin and aldosterone secretion at high salt intakes shifts the top of the renal function curve toward the left, and the increased renin and aldosterone secretion at low salt intakes shifts the bottom of the renal function curve toward the right. In support of these findings, note the relative renin values at different levels of the chronic renal function curve as recorded by DeClue et al (10), illustrating marked reduction of renin secretion at high sodium level and marked increase in renin at low sodium level. Quantitatively, these effects of renin and aldosterone can account for almost half of the total change from the sloping acute curve to the very steep chronic renal function curve. Other factors that may also play roles in this effect are long-term intrarenal hemodynamic changes and natriuretic factors, but the quantitative importance of both of these is yet to be established.

Because of the widely heralded ability of renin to increase the rate of aldosterone secretion, it is tempting to ascribe the steepening of the renal function curve by both renin and aldosterone entirely to changes in aldosterone, with the renin-angiotensin system functioning by changing the rate of aldosterone secretion. However, many different studies in which the aldosterone factor has been controlled show that angiotensin very potently and directly causes the kidneys to retain sodium (18–20); angiotensin blockade causes markedly increased excretion of sodium (21, 22). In fact, the quantitative results of all these studies suggest that the direct effect of angiotensin on the kidneys is several times as potent as the direct effect of aldosterone to cause sodium retention. Therefore, there is accumulating evidence that this direct action of angiotensin on the kidneys is the major mechanism for sodium conservation rather than the effect of the aldosterone.

In summary, because the kidney can excrete extreme amounts of salt with very little rise in arterial pressure, massive salt loading in normal animals and normal humans does not cause hypertension.

Hemodynamics

By the term "short-term salt-loading hypertension" we mean hypertension that occurs in a matter of days following buildup of excess sodium in the body. Then, when the excess sodium is removed, arterial pressure returns to its normal level within another few days. It is not possible to cause such hypertension in most normal animals by acutely loading them with large excesses of salt, but it can be achieved by a combination of salt loading and almost any maneuver that inhibits the kidney's ability to excrete salt—for instance, removing kidney mass.

Figure 2 illustrates the circulatory hemodynamic changes that lead to hypertension at the onset of chronic salt loading. It is a composite diagram from experiments by Langston et al (7), Douglas et al (8), Coleman & Guyton (9), and Manning et al (11), in which approximately 70% of the kidney mass was first removed and then the animal's daily salt intake was increased to approximately five times normal. The sequential changes were the following:

The extracellular fluid volume, blood volume, mean circulatory filling pressure, pressure gradient for venous return, and cardiac output all increased 20–60% during the first few days, but after about two weeks all of these had returned to almost normal, with the exception of the mean circulatory filling pressure, which remained 20% above normal. The "mean circulatory filling pressure" is the pressure measured in the entire circulatory system when the heart stops pumping and the pressures everwhere in the circulation are allowed to equalize. Thus, the mean circulatory filling pressure is a measure of how tightly the circulatory system is filled with blood. And the "pressure gradient for venous return" is the difference between the mean circulatory filling pressure and the right atrial pressure; this difference is a measure of the driving force that causes return of blood from the peripheral circulation to the heart.

The total peripheral resistance actually decreased during the first few days of salt loading, decreasing to about 20% below normal at the end of the first day but then rising to 38% above normal after about two weeks.

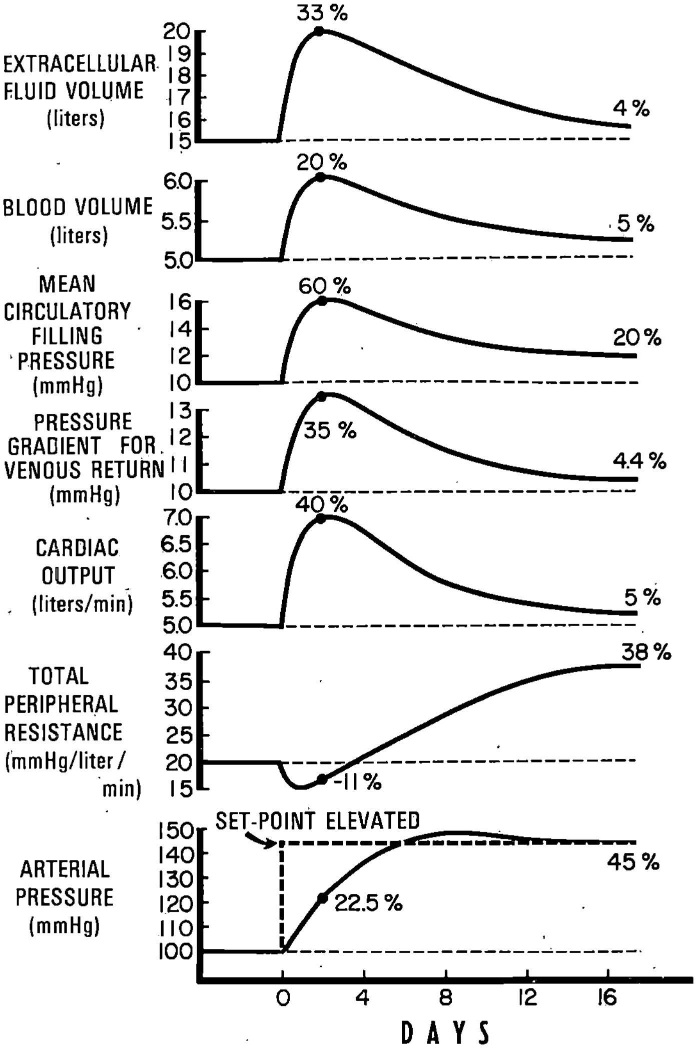


Figure 2 Sequential changes in the important cardiovascular hemodynamic variables in the first few weeks following the onset of short-term salt-loading hypertension. Note especially that the arterial pressure rises ahead of the increase in total peripheral resistance. (Preprinted from Guyton, A. C. 1980. *Arterial Pressure Control and Hypertension,* Philadelphia: Saunders. In press.)

The arterial pressure rose progressively for the first several days and at the end of a week or so leveled out at a new plateau level about 45% above normal.

Now, let us examine these hemodynamic changes. Five of the factors were all related to the increased fluid volume of the body—the extracellular fluid volume, the blood volume, the mean circulatory filling pressure, the pressure gradient for venous return, and the cardiac output. These were transiently but very significantly increased during the first few days of excess salt. It was during these same first few days that the arterial pressure rose most rapidly. However, by the time the arterial pressure had established its steady-state elevated level, all of the factors related to increased volume had returned almost to normal. In fact, none of the factors, with the possible exception of the mean circulatory filling pressure, could reliably be measured as abnormal by that time. Thus, here is an instance of true volume-loading hypertension, and yet once the steady state had been reached it was almost impossible to demonstrate that the volume itself, or other factors directly related to it, had actually increased.

Note that the total peripheral resistance decreased during the first few days of excess salt intake. It was during this time that most of the arterial pressure increase occurred. Therefore, the total peripheral resistance not only did not contribute to the rise in arterial pressure, it actually inhibited the pressure rise. Once the arterial pressure had reached its hypertensive level, the total peripheral resistance, almost as an afterthought, then increased progressively nearly as much as the arterial pressure had increased. During the course of this increase in total peripheral resistance, all of the volume factors returned nearly to normal.

There are many reasons to believe that the changes in total peripheral resistance recorded in Figure 2 were caused in the following manner: The initial decrease in total peripheral resistance was almost certainly caused by the baroreceptor reflex that caused peripheral vasodilatation. But the secondary increase in total peripheral resistance is attributed to the autoregulation mechanism (23, 34). This is the effect of excess blood flow through tissues to cause a progressive increase in local vascular resistance until the flow returns nearly to normal. The increase in resistance in all parts of the body in turn decreases the cardiac output back toward normal. But it also increases the total peripheral resistance. As shown in Figure 2, this increase in resistance is very marked, increasing almost as much as the increase in arterial pressure—but the increase in resistance follows the increase in pressure; it does not precede it.

Therefore, by the time short-term salt-loading hypertension reaches steady state (in only a few weeks), it is a high resistance hypertension, not a high cardiac output hypertension, even though its initiating cause was high volume and high cardiac output.

The Volume Factor

We have already alluded to the very small increases in extracellular fluid volume and blood volume found during the steady-state stage of short-term salt-loading hypertension. Because of this finding, confirmed by many others, it has often been argued that the hypertension in salt-loading states could not be caused by an increase in volume. However, there are several reasons for believing this is not a reasonable conclusion. First, as illustrated in Figure 2, the hypertension initially developed when there were definite increases in both extracellular fluid volume and blood volume, and also increases in all the other hemodynamic factors related to excess volume. Thus, the hypertension became established even before the total peripheral resistance rose. Second, if the autoregulation is even reasonably powerful, a steady-state increase in cardiac output of as little as 5% could easily increase the total peripheral resistance 38% as illustrated in Figure 2. This would require an autoregulation feedback "gain" of approximately 7, a value that has actually been measured under some conditions in animals during the course of whole-body autoregulation experiments (25, 26). Third, several different experiments have examined whether or not excess body salt can cause hypertension in the absence of increased volume. For instance, Norman et al (27) increased the extracellular fluid sodium concentration approximately 20% but at the same time used a combination of hemodialysis and water restriction to keep the extracellular fluid volume from increasing. They found no significant increase in arterial pressure. When they allowed the animals to drink sufficient water, the extracellular fluid volume expanded immediately, and hypertension developed in the next few days. In another series of experiments, Manning et al (28) used a combination of decreased kidney mass, infusion of ADH, and infusion of fluid (*a*) to reduce the sodium concentration in the plasma of dogs to approximately 125 mEq per liter, and (*b*) to increase the blood volume above normal. Despite the very low sodium concentration, hypertension still developed during the high volume state.

Thus, all of the experimental evidence thus far indicates that salt causes hypertension not because of a direct effect of the salt itself on the circulation but because of its propensity to increase the extracellular fluid and blood volumes.

The Kidney Factor

It must be reemphasized that salt loading alone will not cause short-term salt-loading hypertension. This requires a kidney factor as well, because, as has already been discussed, normal kidneys will simply excrete the excess salt with very little rise in pressure. But several different types of abnormal renal function can provide this required kidney factor. Figure 3 graphically illustrates the roles played by three different kidney abnormalities, and

P 24,842 2195 c.1

displays the normal renal function curve, which is the same as the normal curve shown in Figure 1, for comparison. The three abnormal renal function curves represent (*a*) chronic infusion of aldosterone into a dog at a rate of 100 μg/day (J. DeClue and A. C. Guyton, unpublished observations); (*b*) chronic infusion of angiotensin into a dog at 5 ng/kg/min (10); and (*c*) chronic reduction of the kidney mass of a dog to 30% of normal [drawn from data in (8) and (9)].

Also shown in Figure 3 are two levels of sodium intake: normal and five times normal. In the steady state, the urinary output of sodium must exactly equal the daily sodium intake. Therefore, the points at which the sodium intake levels cross the respective curves dictate the pressure levels at which the long-term steady-state arterial pressure will stabilize in each set of conditions. Note that when the sodium intake is normal, even the abnormal renal function curves cross the intake level at pressures no greater than 114

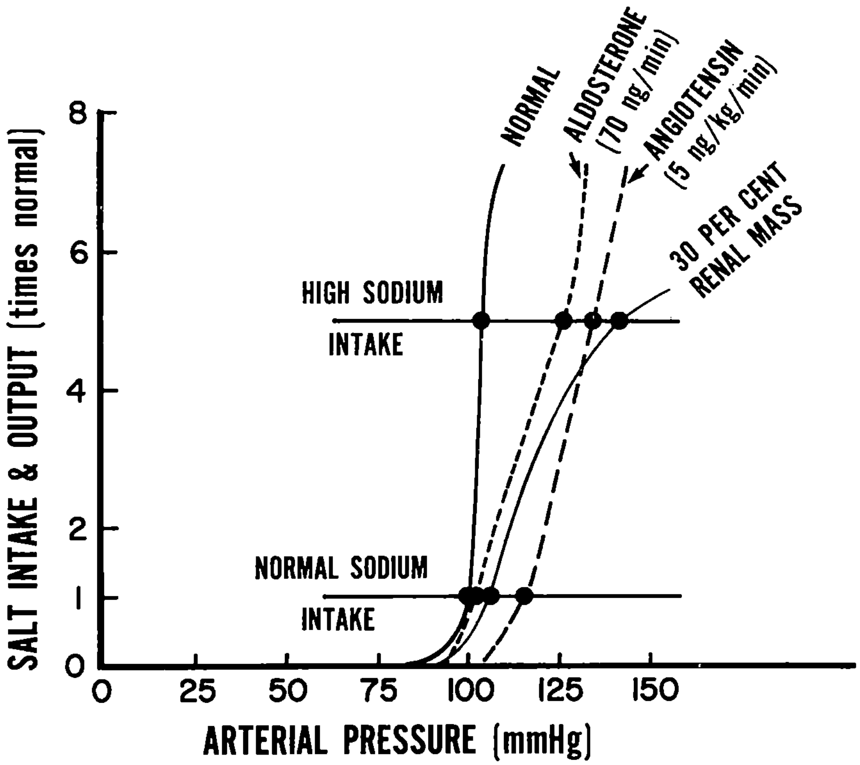


Figure 3 This figure shows the normal renal function curve and three abnormal renal function curves caused respectively by aldosterone infusion, angiotensin infusion, and decrease in kidney mass to 30% of normal.

mm Hg. Thus, the levels of aldosterone and angiotensin infused in these experiments is not sufficient to cause significant hypertension so long as the sodium intake is normal. Also, 70% removal of kidney mass does not cause hypertension.

When the sodium intake is increased, though this causes essentially no increase in blood pressure in the normal animal, nevertheless, in all the animals with the abnormal renal conditions the blood pressure rises to hypertensive levels 25–42 mm Hg above normal.

In summary, the normal animal will not develop hypertension even with tremendous salt loads that last for several weeks. On the other hand, any abnormality of the kidneys that alters the renal function curve so that it slopes far to the right will create a salt-sensitive state in the animal so that hypertension will develop with high salt loads. That is, the kidney factor is essential for the development of short-term salt-loading hypertension.

LONG-TERM SALT LOADING AND HYPERTENSION

We now consider the crucial question, what causes long-term salt-loading hypertension? We have presented the details of the much better understood short-term salt-loading hypertension as a basis for discussing this problem. Because one can almost never show evidence of excess fluid volume or increased cardiac output in long-term salt-loading hypertension, most research workers have sought evidence that the salt directly increases the total peripheral resistance (29). However, we have already seen that within a few weeks after the onset of short-term salt-loading hypertension, one cannot reliably measure any increased body fluid volume or cardiac output. Even so, all the evidence still indicates that this type of hypertension initially arises from excess fluid volume, and that in the chronic stage the autoregulation mechanism causes the fluid volumes and cardiac output to return to normal and the total peripheral resistance to increase. Thus, the increase in total peripheral resistance is secondary to the hypertension, not its primary cause.

However, there is an even stronger reason for believing that long-term salt-loading hypertension does not result from any direct effect of salt to cause a primary increase in total peripheral resistance: Other factors that do cause primary changes in the long-term level of total peripheral resistance do not alter the arterial pressure. For instance, in patients with very large A-V fistulas, closing these fistulas does not raise the pressure significantly, even though the total peripheral resistance is often increased more than 100% (30). Also, in patients with all four limbs removed, the arterial pressure still remains normal. Conversely, in beriberi, hyperthyroidism, and anemia, in all of which the total peripheral resistance is markedly decreased, the arterial pressure does not fall below normal as long as the heart does not fail.

The Kidney Factor

Since long-term salt-loading hypertension cannot be ascribed to a primary increase in total peripheral resistance, we must look elsewhere. Short-term salt-loading hypertension cannot occur without a kidney abnormality to cause salt retention. But, as a mental exercise, let us consider for a moment what would happen if kidney function were entirely normal and hypertension did occur because of some factor that is not kidney related. Referring to the normal chronic renal function curve shown in Figures 1, 3, and 4, one can see that the elevated pressure would immediately trigger extreme excess of sodium output. This would continue until the blood pressure fell to a point where it would lie on the normal renal function curve. Thus, hypertension caused by any factor that does not cause a simultaneous change in kidney function will be self-correcting. This logic implies that the kidney must become a salt-retaining organ; otherwise the state of long-term salt-loading hypertension cannot occur. This is illustrated in Figure 4,

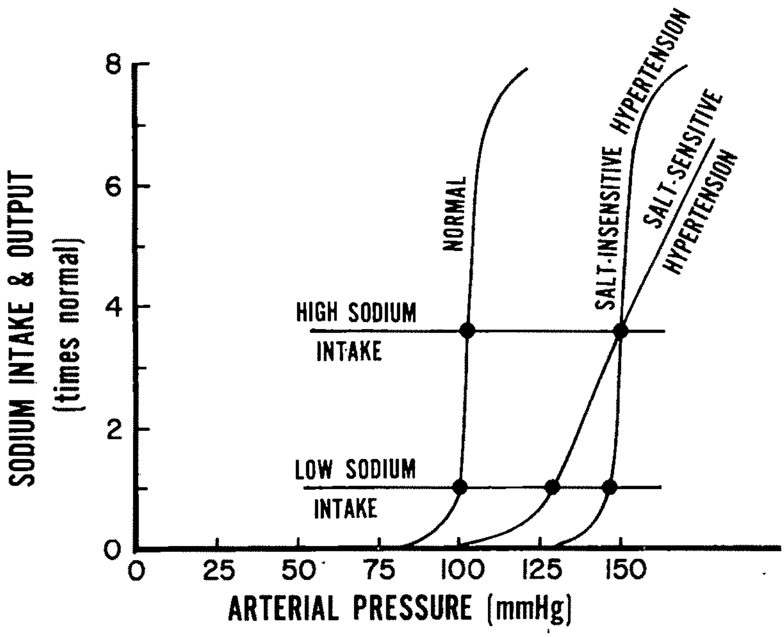


Figure 4 Postulated shift of the renal function curve from normal pressure levels to hypertensive levels.

which shows the normal renal function curve and two hypertensive renal function curves, one for salt-insensitive hypertension and the other for salt-sensitive hypertension. Similar salt-sensitive and salt-insensitive curves have been recorded by Bartter and his colleagues in different patients with essential hypertension (31). The other evidence presented thus far in this paper suggests that such curves also occur in persons with long-term salt-loading hypertension.

Yet, what is it about long-term salt loading that can cause the renal function curve to become a hypertensive type of curve rather than the normal curve? This question cannot be answered at present. However, bits of evidence suggest that excess salt intake over months and years causes actual damage to the kidneys that eventuates in the hypertensive types of renal function curve. This was convincingly demonstrated by Dahl and his colleagues (32–34) in a genetic strain of rats whose kidneys were susceptible to damage when they ingested excessive amounts of salt. The rats developed permanent hypertension that remained even when the salt load was reduced to normal. Meneely et al (35, 36), Barsanto et al (37), Mann et al (38), and others have all shown that excess salt loading also accelerates the development of hypertension in other strains of rats and that the hypertension will only partially abate when the excess salt intake is stopped. In many of these studies, actual pathologic lesions have been demonstrated in the kidney tissue.

In summary, it is likely that long-term salt-loading hypertension cannot occur without the development of a kidney abnormality. And there have also been experimental demonstrations that excessive salt-loading can, under appropriate conditions, cause permanent changes in renal function, even to the extent of causing intrarenal lesions. Therefore, it would also be logical to believe that the long-term development of elevated blood pressure in persons who ingest increased amounts of salt probably results from progressive pathological change in the kidneys themselves.

Literature Cited

1. Prior, I. A. M., Evans, R. S., Harvey, H. P. B., Davidson, F., Lindsey, M. 1968. Sodium intake and blood pressure in two Polynesian populations. *N. Engl. J. Med.* 279:515–20
2. Beaglehole, R., Salmond, C. E., Hooper, A., Huntsman, J., Stanhope, J. M., Cassel, J. C., Prior, I. A. M. 1977. Blood pressure and social interaction in Tokelauan migrants in New Zealand. *J. Cron. Dis.* 30:803–12
3. Page, L. B., Damon, A., Moellering, R. C. 1974. Antecedents of cardiovascular diseases in six Solomon Island populations. *Circulation* 49:1132–46
4. Oliver, W. J., Cohen, E. L., Neel J. V. 1975. Blood pressure, sodium intake, and sodium related hormones in the Yanomano Indians. A "no-salt" culture. *Circulation* 52:146–51
5. Dahl, L. K. 1961. Possible role of chronic excess salt consumption in the pathogenesis of essential hypertension. *Am. J. Cardiol.* 8:571–75
6. Sasaki, N. 1962. High blood pressure and the salt intake of the Japanese. *Jpn. Heart J.* 3:313–24
7. Langston, J. B., Guyton, A. C., Douglas, B. H., Dorsett, P. E. 1963. Effect of changes in salt intake on arterial pressure and renal function in nephrectomized dogs. *Circ. Res.* 12:508–13
8. Douglas, B. H., Guyton, A. C., Langston, J. B., Bishop, V. S. 1964. Hypertension caused by salt loading. II. Fluid volume and tissue pressure changes. *Am. J. Physiol.* 207:669–71
9. Coleman, T. G., Guyton, A. C. 1969. Hypertension caused by salt loading in the dog. III. Onset transients of cardiac output and other circulatory variables. *Circ. Res.* 25:152–60
10. DeClue, J. W., Guyton, A. C., Cowley, A. W. Jr., Coleman, T. G., Norman, R. A. Jr., McCaa, R. E. 1978. Subpressor angiotensin infusion, renal sodium handling, and salt-induced hypertension in the dog. *Circ. Res.* 43:503–12
11. Manning, R. D. Jr., Coleman, T. G., Guyton, A. C., Norman, R. A. Jr., McCaa, R. E. 1979. Essential role of mean circulatory filling pressure in salt-induced hypertension. *Am. J. Physiol.* 236:R40–R47
12. Murray, R. H., Luft, F. C., Block R., Weyman, A. E. 1978. Blood pressure response to extremes of sodium intake in normal man. *Proc. Soc. Exp. Biol. Med.* 159:432–36
13. Selkurt, E. E., Hall, P. W., Spencer, M. P. 1949. Influence of graded arterial pressure decrement on renal clearance of creatinine, *p*-amino hippurate and sodium. *Am. J. Physiol.* 159:369–84
14. Thompson, D. D., Pitts, R. F. 1952. Effects of alterations of renal arterial pressure on sodium and water excretion. *Am. J. Physiol.* 168:490–99
15. Thurau, K., Deetzen, P. 1962. Diuresis in arterial pressure increases. *Pfluegers Arch.* 274:567–80
16. Young, D. B., Guyton, A. C. 1977. Steady-state aldosterone dose-response relationships. *Circ. Res.* 40:138–42
17. Pan, Y. J. 1979. *Experimental aldosterone hypertension.* PhD dissertation, Univ. Miss. Med. Cent., Jackson, Miss.
18. Yeyati, N. L., Taquini, C. M., Etchev, erry, J. C. 1972. A possible role of angiotensin in the regulation of glomerular balance for sodium. *Medicina* 32: (Suppl. I), pp. 68–73
19. Waugh, W. H. 1972. Angiotensin II: Local renal effects of physiological increments in concentration. *Can. J. Physiol. Pharmacol.* 50:711–16
20. Fagard, R. H., Cowley, A. W. Jr., Navar, L. G., Langford, H. G., Guyton, A. C. 1976. Renal responses to slight elevations of renal arterial plasma angiotensin II concentration in dogs. *Clin. Exp. Pharmacol. Physiol.* 3:19–26
21. Hall, J. E., Guyton, A. C., Trippodo, N. C., Lohmeier, T. E., McCaa, R. E., Cowley, A. W. Jr. 1977. Intrarenal control of electrolyte excretion by angiotensin II. *Am. J. Physiol.* 252:F538–F544
22. Lohmeier, T. E., Cowley, A. W. Jr., Trippodo, N. C., Hall, J. E., Guyton, A. C. 1977. Effects of endogenous angiotensin II on renal sodium excretion and renal hemodynamics. *Am. J. Physiol.* 233:F388–F395.
23. Guyton, A. C., Coleman, T. G. 1967. Long-term regulation of the circulation; Interrelationships with body fluid volumes. In *Physical Bases of Circulatory Transport: Regulation and Exchange,* ed. E. B. Reeve, A. C. Guyton, pp. 179–201. Philadelphia: Saunders.
24. Guyton, A. C., Coleman, T. G., Cowley, A. W. Jr., Manning, R. D. Jr., Norman, R. A. Jr., Ferguson, J. D. 1974. A systems analysis approach to understanding long-range arterial blood pressure control and hypertension. *Circ. Res.* 35:159–76
25. Granger, H. J., Guyton, A. C. 1969. Autoregulation of the total systemic circulation following destruction of the central nervous system in the dog. *Circ. Res.* 25:379–88

26. Shepherd, A. P., Granger, H. J., Smith, E. E., Guyton, A. C. 1973. Local control of tissue oxygen delivery and its contribution to the regulation of cardiac output. *Am. J. Physiol.* 225:747–55
27. Norman, R. A. Jr., Coleman, T. G., Wiley, T. L. Jr., Manning, R. D. Jr. Guyton, A. C. 1975. Separate roles of sodium ion concentration and fluid volumes in salt-loading hypertension in sheep. *Am. J. Physiol.* 229:1068–72
28. Manning, R. D. Jr., Guyton, A. C., Coleman, T. G., McCaa, R. E. 1979. Hypertension in dogs during antidiuretic hormone and hypotonic saline infusion. *Am. J. Physiol.* 236:H314–H322
29. Friedman, S. M., Friedman, C. L. 1963. Effects of ions on vascular smooth muscle. In *Handbook of Physiology-Circulation,* ed. W. F., Hamilton, P. Dow, 2:1135–1166. Washington, D.C: Am. Physiol. Soc.
30. Warren, J. V., Nicherson, J. L, Elkin, D. C. 1951. The cardiac output in patients with arteriovenous fistulas. *J. Clin. Invest.* 30:210–14
31. Kawasaki, T., Delea, C. S., Bartter, F. C., Smith, H. 1978. The effect of high-sodium and low-sodium intakes on blood pressure and other related variables in human subjects with idiopathic hypertension. *Am. J. Med.* 64:193–98
32. Dahl, L. K. Heine, M., Tassinari, L. 1963. Effects of chronic excess salt ingestion: Role of genetic factors in both DOCA-salt and renal hypertension. *J. Exp. Med.* 118:605–17
33. Dahl, L. K., Knudsen, K. D., Heine, M., Leitl, G. 1968. Effects of chronic excess salt ingestion. Modification of experimental hypertension in the rat by variations in the diet. *Circ. Res.* 22:11–18
34. Dahl, L. K., Leitl, G., Heine, M. 1972. Influence of dietary potassium and sodium/potassium molar ratios on the development of salt hypertension. *J. Exp. Med.* 136:318–30
35. Meneely, G. R., Tucker, R. G., Darby, W. J., Auerbach, S. H. 1953. Chronic sodium chloride toxicity in albino rat. II. Occurrence of hypertension and syndrome of edema and renal failure. *J. Exp. Med.* 98:71–80
36. Meneely, G. R., Ball, C. O. T., Youmans, J. B. 1957. Chronic sodium chloride toxicity: The protective effect of added potassium chloride. *Ann. Intern. Med.* 47:263–73
37. Barsanti, J. A., Pillsbury, H. R. C. III, Freis, E. D. 1971. Enhanced salt toxicity in the spontaneously hypertensive rat. *Proc. Soc. Exp. Biol. Med.* 136:565–68
38. Mann, J. F. E., Dietz, R., Rascher, W., Schomig, A., Luth, J. B., Gross, R. 1978. Kidney function during the development of genetic hypertension in rats. *Proc. 8th World Cong. Cardiol., Tokyo, Abstr.* Vol. II, p. 928

Ann. Rev. Med. 1980. 31:29–34
Copyright © 1980 by Annual Reviews Inc. All rights reserved

HEMOGLOBIN A_{Ic} AND DIABETES MELLITUS

◆7340

Ronald J. Koenig, Ph.D., M.D.

Peter B. Brigham Hospital, Boston, Massachusetts 02115

Anthony Cerami, Ph.D.

Rockefeller University, New York, NY 10021

Hemoglobin (Hb) A_{Ic} is a minor red cell constituent that comprises 5% of the total Hb in normal individuals but up to 15% in patients with diabetes mellitus (1). It may be assayed by ion exchange chromatography (1–3), isoelectric focusing (4), radioimmunoassay (5), or colorimetry (6). Current data suggest that the quality of diabetic control can best be assessed by monitoring Hb A_{Ic} levels, and that the synthesis of Hb A_{Ic} may represent a unifying model reaction that underlies the pathogenesis of many of the sequelae of chronic diabetes. This review summarizes the biochemical and clinical implications of Hb A_{Ic}.

STRUCTURE AND BIOSYNTHESIS

Hemoglobin A_{Ic} and the major adult hemoglobin, Hb A, have identical amino acid sequences (7). The unique structural feature of Hb A_{Ic} is the presence of 1-deoxy fructose linked through carbon number 1 to the NH_2-terminus of the β chain (8, 9). This apparently derives nonenzymatically from glucose forming a Schiff base with the NH_2-terminal valine and subsequently undergoing an Amadori rearrangement (10). [For a review of whether glucose-6-phosphate may be the reacting molecule in Hb A_{Ic} formation, see (11)].

The biosynthesis of Hb A_{Ic} has been investigated in wild-type and diabetic mice (12) as well as in nondiabetic humans (13). In both species Hb A_{Ic} is synthesized at a slow rate in a nearly irreversible fashion throughout

0066-4219/80/0401-0029$01.00

the life of the mature erythrocyte. Since erythrocytes are not capable of protein synthesis, this continual Hb A_{Ic} formation must result from post-synthetic glycosylation of already existing Hb A. Furthermore, when red cells from wild-type mice were injected into diabetic recipients, their rate of Hb A_{Ic} synthesis increased 2.7-fold.

As these studies would predict, older red cells contain the most Hb A_{Ic} (14), and patients with hemolytic anemias have decreased Hb A_{Ic} levels (13). A steady-state level of Hb A_{Ic} is achieved through the replacement of old red cells laden with Hb A_{Ic} by young erythrocytes relatively poor in the Hb. The actual Hb A_{Ic} concentration is a function of red cell half-life and rate of Hb A_{Ic} synthesis, which in turn is determined by the blood glucose concentration.

CARBOHYDRATE CONTROL

Hemoglobin A_{Ic} is synthesized in a slow, nearly irreversible reaction throughout the life of the red cell; the rate of synthesis is a function of blood glucose concentration. A severely hyperglycemic subject should therefore have a very high Hb A_{Ic} concentration. If one abruptly lowered that subject's blood glucose level, the rate of Hb A_{Ic} synthesis would also fall abruptly. However, the peripheral blood level of Hb A_{Ic} would not show an immediate decrease, because erythrocytes that had already synthesized large amounts of Hb A_{Ic} would remain in the circulation until their 120-day lifespans expired. Thus, a single Hb A_{Ic} measurement should reflect a patient's mean blood glucose concentration over the previous two to three months. This hypothesis has been substantiated by the following investigations.

In newly diabetic mice, Hb A_{Ic} levels do not increase until hyperglycemia has been present for three weeks (12). In human diabetics, Hb A_{Ic} shows a weak but significant correlation with fasting blood sugar and a stronger correlation with response to an oral glucose tolerance test (15).

When a group of poorly controlled diabetics were hospitalized to achieve strict carbohydrate control (16), the decrease in Hb A_{Ic} lagged three weeks behind the institution of rigid carbohydrate control. An excellent correlation was obtained between Hb A_{Ic} and estimated mean daily blood sugar levels both prior to and during strict control.

In a study of insulin-dependent diabetics (17), Hb A_{Ic} was found to correlate with the mean urinary glucose for the preceding eight weeks, but not with occasional checks of fasting or postprandial blood sugar. This implies that the level of this Hb correlates with long-term, rather than spot checks of, diabetic control. A correlation has also been reported (18) between Hb A_{Ic} and urinary glucose measured several times over the previous three months. Similar data have been obtained by other investigators (19, 20).

In a study of diabetic pregnancies (21), third trimester maternal Hb A_{Ic} was shown to correlate with infant birth weight. Since Hb A_{Ic} levels should remain elevated for weeks following recovery from gestational diabetes, mothers of large-for-dates babies or babies with birth defects could be screened postpartum for gestational diabetes.

Recent preliminary data (22) suggest the formation of Hb A_{Ic} may be more reversible than initially thought. Whether or not this will be significant enough to influence the time span over which Hb A_{Ic} integrates blood glucose levels has yet to be determined.

In summary, a single Hb A_{Ic} measurement reflects the patient's mean blood sugar concentration for the preceding weeks to months. Infrequent quantification of Hb A_{Ic} on an outpatient basis can give an objective, accurate assessment of long-term diabetic control. In addition, Hb A_{Ic} measured serially over time in a prospective fashion can be used to define the role of hyperglycemia in the development of the sequelae of chronic diabetes.

COMPLICATIONS OF DIABETES

The synthesis of Hb A_{Ic} also may represent a model reaction to explain the biochemical basis for many of the long-term sequelae of diabetes. The tissues that suffer the most noticeable dysfunctions in diabetes (e.g. retina, lens, peripheral nerve, kidney) appear to be insulin independent for glucose uptake (23). Perhaps in diabetes the intracellular proteins of these tissues undergo excess nonenzymatic glycosylation, analogous to that seen within red cells. Such glycosylation might alter the enzymatic activity, solubility, antigenicity, or other function of the protein and thereby result in the observed clinical dysfunction. It should be noted that although the glycosylation of hemoglobin results in a slightly increased oxygen affinity (24), this is not likely to be clinically significant since mutant hemoglobins with much greater oxygen affinities are not associated with clinical disease (25).

The role of nonenzymatic glycosylation in diabetic cataract formation is currently being investigated (26). Bovine and rat lens proteins undergo nonenzymatic glycosylation of lysine ϵ amino groups in the presence of a high glucose environment. This glycosylation imparts to the proteins an increased susceptibility to form high molecular weight aggregates, which are large enough to diffract light and result in opacities. The accumulation of sorbitol has also been proposed as a mechanism of diabetic cataractogenesis (27), and the relative roles of these two mechanisms are currently unclear.

The serum proteins albumin (28) and insulin (29) also undergo nonenzymatic glycosylation of lysine ϵ amino groups. Whether or not these modifications have biological importance remains to be determined.

Cyanate is an experimental antisickling agent that forms covalent bonds with amino groups, including the NH_2-termini of hemoglobin chains. Thus, cyanate modifies by carbamylation the same sites that glucose modifies by nonenzymatic glycosylation. The primary toxicities of chronic cyanate ingestion are peripheral neuropathy (30) and cataracts (31), which are histologically similar to the respective lesions seen in diabetes. The cyanate-induced lesions are potentially reversible (31, 32), which implies that the lesions of diabetes may be reversible or at least controllable if euglycemia can be achieved. The rate of reversibility would be a function of protein turnover. In addition, nonenzymatic glycosylation may be somewhat reversible, allowing the glyco groups to elute slowly from proteins with long biological half-lives. Indeed, certain sequelae of diabetes, e.g. platelet hyperaggregability, are reversed when carbohydrate is rigidly controlled (33).

Nonenzymatic glycosylation is not unique to the diabetic state; rather, the higher the blood sugar level the faster the reaction. Thus, the complications seen in diabetes might be expected to develop in the general population, albeit at a slower rate. Senile cataracts, for example, might also be due to nonenzymatic glycosylation. In this sense, diabetes would represent premature aging.

SUMMARY

Hemoglobin A_{Ic} is produced when glucose reacts nonenzymatically with the NH_2-termini of Hb A β chains and then undergoes an Amadori rearrangement. The concentration of Hb A_{Ic} measured at any given time reflects a patient's mean blood glucose level for the preceding weeks to months. Infrequent measurements of Hb A_{Ic} can therefore be used to assess long-term carbohydrate control in outpatient diabetics. In addition, the synthesis of Hb A_{Ic} may represent a model reaction to explain the pathogenesis of many of the sequelae of chronic diabetes. Nonenzymatic glycosylation reactions may also underlie some of the changes ascribed to normal aging.

NOTE ADDED IN PROOF Hemoglobin A also undergoes nonenzymatic glycosylation at its α chain NH_2-termini and lysine ϵ amino groups (34). These glycosylated species, as well as Hb A_{Ic}, are present in elevated concentrations in diabetic patients (35).

Literature Cited

1. Trivelli, L. A., Ranney, H. M., Lai, H.-T. 1971. Hemoglobin components in patients with diabetes mellitus. *N. Engl. J. Med.* 284:353–57
2. Cole, R. A., Soeldner, J. S., Dunn, P. J., Bunn, H. F. 1978. A rapid method for the determination of glycosylated hemoglobins using high pressure liquid chromatography. *Metabolism* 27:289–301
3. Davis, R. E., Nichol, D. J. 1978. A rapid simplified method for routine measurement of glycosylated hemoglobin. *Lancet* 2:350–51
4. Basset, P., Beuzard, Y., Gavel, M. E., Rosa, J. 1978. Isoelectric focusing of human hemoglobin. *Blood* 51:977–82
5. Javid, J., Pettis, P. K., Koenig, R. J., Cerami, A. 1978. Immunologic characterization and quantification of haemoglobin A_{Ic}. *Br. J. Haematol.* 38:329–37
6. Fluckiger, R., Winterhalter, K. H. 1976. In vitro synthesis of hemoglobin A_{Ic}. *FEBS Lett.* 71:356–60
7. Holmquist, W. R., Schroeder, W. A. 1966. A new N-terminal blocking group involving a Schiff base in hemoglobin A_{Ic}. *Biochemistry* 5:2489–2503
8. Bunn, H. F., Haney, D. N., Gabbay, K. H., Gallop, P. M. 1975. Further identification of the nature and linkage of the carbohydrate in hemoglobin A_{Ic}. *Biochem. Biophys. Res. Commun.* 67:103–9
9. Koenig, R. J., Blobstein, S. H., Cerami, A. 1977. Structure of carbohydrate of hemoglobin A_{Ic}. *J. Biol. Chem.* 252:2992–97
10. Hodge, J. E. 1955. The Amadori rearrangement. *Adv. Carbohydr. Chem.* 10:169–205
11. Koenig, R. J., Cerami, A. 1979. Glycohemoglobins in the adult erythrocyte. *Curr. Top. Hematol.* In press
12. Koenig, R. J., Cerami, A. 1975. Synthesis of hemoglobin A_{Ic} in normal and diabetic mice: potential model of basement membrane thickening. *Proc. Natl. Acad. Sci. USA* 72:3687–91
13. Bunn, H. F., Haney, D. N., Kamin, S., Gabbay, K. H., Gallop, P. M. 1976. The biosynthesis of human hemoglobin A_{Ic}. *J. Clin. Invest.* 57:1652–59
14. Fitzgibbons, J. F., Koler, R. D., Jones, R. T. 1976. Red cell age-related changes of hemoglobins A_{Ia+b} and A_{Ic} in normal and diabetic subjects. *J. Clin. Invest.* 58:820–24
15. Koenig, R. J., Peterson, C. M., Kilo, C., Cerami, A., Williamson, J. R. 1976. Hemoglobin A_{Ic} as an indicator of the degree of glucose intolerance in diabetes. *Diabetes* 25:230–32
16. Koenig, R. J., Peterson, C. M., Jones, R. L., Saudek, C., Lehrman, M., Cerami, A. 1976. Correlation of glucose regulation and hemoglobin A_{Ic} in diabetes mellitus. *N. Engl. J. Med.* 295:417–20
17. Lanoe, R., Soria, J., Thibult, N., Soria, C., Eschwege, E., Tchobroutsky, C. 1977. Glycosylated hemoglobin concentrations and clinical test results in insulin-dependent diabetes. *Lancet* 2:1156–57
18. Gabbay, K. H., Hasty, K., Breslow, J. L., Ellison, R. C., Bunn, H. F., Gallop, P. M. 1977. Glycosylated hemoglobins and long-term blood glucose control in diabetes mellitus. *J. Clin. Endocrinol. Metab.* 44:859–64
19. Gonen, B., Rubenstein, A. H., Rochman, H., Tanega, S. P., Horowitz, D. L. 1977. Haemoglobin A_I: an indicator of the metabolic control of diabetic patients. *Lancet* 2:734–36
20. Tze, W. J., Thompson, K. H., Leichter, J. 1978. Hemoglobin A_{Ic}—an indicator of diabetic control. *J. Pediatr.* 93:13–16
21. Widness, J. A., Schwartz, J. C., Thompson, D., King, K. C., Kahn, C. B., Oh, W., Schwartz, R. 1978. Glycohemoglobin (Hb A_{Ic}): a predictor of birth weight in infants of diabetic mothers. *J. Pediatr.* 92:8–12
22. Bolli, G., Cartechini, M. G., Compagnucci, P., Massi-Benedetti, M., Santeusanio, F., Paolo, B. 1979. Fast glycosylation of haemoglobin. *Lancet* 1:1143–44
23. Spiro, R. G. 1976. Search for a biochemical basis for diabetic microangiopathy. *Diabetologia* 12:1–14
24. McDonald, M. J., Bleichman, M., Bunn, H. F., Noble, R. W. 1979. Functional properties of the glycosylated minor components of human adult hemoglobin. *J. Biol. Chem.* 254:702–7
25. Charache, S. 1974. Haemoglobins with altered oxygen affinity. *Clin. Haematol.* 3:357–83
26. Stevens, V. J., Rouzer, C. A., Monnier, V. M., Cerami, A. 1978. Diabetic cataract formation: potential role of glycosylation of lens crystallins. *Proc. Natl. Acad. Sci. USA* 75:2918–22
27. Kinoshita, J. H. 1974. Mechanisms initiating cataract formation. *Invest. Ophthalmol.* 13:713–24

28. Day, J. F., Thorpe, S. R., Baynes, J. W. 1979. Nonenzymatically glucosylated albumin. *J. Biol. Chem.* 254:595–97
29. Dolhofer, R., Wieland, O. H. 1979. Preparation and biological properties of glycosylated insulin. *FEBS Lett.* 100:133–36
30. Tellez-Nagel, I., Koethals, J. K., Vlassara, H., Cerami, A. 1977. An ultrastructural study of chronic sodium cyanate-induced neuropathy. *J. Neuropath. Exp. Neurol.* 36:351–62
31. Nicholson, D. H., Harkness, D. R., Benson, W. E. 1976. Cyanate-induced cataracts in patients with sickle-cell hemoglobinopathies. *Arch. Ophthalmol.* 94:927–30
32. Peterson, C. M., Tsairis, P., Ohnishi, A., Yang, S. L., Grady, R. W., Cerami, A., Dyck, P. J. 1974. Sodium cyanate–induced neuropathy in patients with sickle cell disease. *Ann. Intern. Med.* 81:852–58
33. Peterson, C. M., Jones, R. L., Koenig, R. J., Melvin, E. T., Lehrman, M. L. 1977. Reversible hematologic sequelae of diabetes mellitus. *Ann. Intern. Med.* 86:425–29
34. Bunn, H. F., Shapiro, R., McManus, M., Garrick, L., McDonald, M. J., Gallop, P. M., Gabbay, K. H. 1979. Structural heterogeneity of human hemoglobin A due to nonenzymatic glycosylation. *J. Biol. Chem.* 254:3892–98
35. Gabbay, K. H., Sosenko, J. M., Banuchi, G. A., Mininsohn, M. J., Fluckiger, R. 1979. *Diabetes* 28:337–40

Ann. Rev. Med. 1980. 31:35–49
Copyright © 1980 by Annual Reviews Inc. All rights reserved

CONTROL OF ACID EXCRETION BY THE KIDNEY

❖7341

Richard L. Tannen, M.D.

Department of Internal Medicine, Nephrology Division,
University of Michigan Medical School; and Veterans
Administration Hospital, Ann Arbor, Michigan 48109

INTRODUCTION

Under normal conditions the renal regulation of acid excretion is closely coupled to hydrogen ion intake, so that plasma bicarbonate concentration is maintained at stable levels despite wide variations in dietary acid ingestion. With larger hydrogen ion loads resulting from altered intake or production, or with abnormalities in CO_2 elimination by the lungs (respiratory acid-base disorders), renal acid excretion responds in a fashion that tends to defend the maintenance of a normal systemic pH.

The renal regulation of hydrogen ion homeostasis can be viewed in terms of two distinct, yet intimately linked, processes—hydrogen ion (acid) excretion and secretion. Acid excretion provides a renal mechanism to modify the body alkali stores. Quantitatively, acid excretion is equivalent to the sum of urinary titratable acid and ammonium, minus the bicarbonate lost in the urine. Hydrogen ion secretion includes the quantity of filtered bicarbonate reabsorbed by the nephron in addition to the amount of acid excreted. Functionally, it determines the plasma bicarbonate concentration that can be sustained. Under normal conditions more than 98% of hydrogen ion secretion is expended in bicarbonate reabsorption, with only a minor fraction accounting for acid excretion. Despite this quantitative discrepancy between secretion and excretion, the processes are integrated so that acid excretion can be meticulously controlled.

This review concentrates specifically on the regulation of acid excretion. The renal tubular events that may modify hydrogen ion excretion and the signal(s) that feeds back to control renal acid excretion are considered. In addition, a model for the regulation of acid excretion is speculated upon.

0066-4219/80/0401-0035$01.00

HYDROGEN ION SECRETION BY THE NEPHRON

In order to focus on the specific tubular events that might be crucial for the regulation of acid excretion, current concepts regarding hydrogen ion secretion and excretion by the nephron are summarized briefly. This topic has been the subject of several excellent reviews (1–6) including a recent succinct discussion in this publication (7).

Renal Tubule

The specific events at a tubular level concerned with hydrogen ion excretion have been elucidated by micropuncture experiments in the intact animal (largely rats and dogs) (1, 3, 4, 6, 8) and by in vitro perfusion of isolated segments of rabbit nephron (9–16). Micropuncture has provided data from the accessible portions of the superficial cortical nephrons of anesthetized animals, while perfusion experiments have supplemented these findings and also provided information regarding tubular segments inaccessible to micropuncture in vivo. By utilizing these observations, a tentative scheme can be proposed for hydrogen ion handling along the nephron.

Approximately 80% of filtered bicarbonate is reabsorbed by the last portion of the superficial proximal convoluted tubule accessible to micropuncture. Recent evidence, which indicates that proximal tubular bicarbonate reabsorption can be completely eradicated by carbonic anhydrase inhibition (10, 11, 17), suggests that this process is entirely dependent on the enzyme carbonic anhydrase. Cellular carbonic anhydrase provides hydrogen ions for secretion by catalyzing their formation from CO_2; enzyme present on the luminal border minimizes the decrease in tubular fluid pH by accelerating the conversion of carbonic acid to CO_2 and H_2O. The activity of the enzyme at both these sites appears to be crucial for normal bicarbonate reabsorption by the proximal convoluted tubule (18). However, carbonic anhydrase is present in vast excess, so its activity is unlikely to be rate limiting for hydrogen ion secretion under normal circumstances (17).

The virtually complete inhibition of bicarbonate reabsorption by carbonic anhydrase inhibition supports the contention that bicarbonate reabsorption by the proximal tubule is not direct; rather it is largely the result of hydrogen ion secretion. The evidence for the actual presence of hydrogen ion secretion is primarily the demonstration of a "disequilibrium pH" under conditions of carbonic anhydrase inhibition (19, 20).[1] This suggests that carbonic acid has accumulated in the tubular fluid, as a result of the reaction between secreted hydrogen ion and bicarbonate. However, the demonstration of a disequilibrium pH was based on the assumption that the tubular fluid pCO_2 was approximately 40 mm Hg. In view of the recent measurements demonstrating that the pCO_2 in both tubular fluid and peritubular capillary blood is substantially greater than arterial pCO_2 (i.e. 40 mm Hg), this issue requires reexamination (21).

---

[1] A "disequilibrium pH" refers to a lower pH of tubular fluid in situ than in fluid removed from the tubule and equilibrated to a similar pCO_2. It implies that the dehydration of carbonic acid to CO_2 does not achieve equilibrium in situ, i.e. there is excess H_2CO_3 in relation to CO_2.

Tubular fluid bicarbonate concentration is approximately equivalent in the early distal tubule and the last accessible portion of the proximal convoluted tubule. Since water is extracted between these two sites, bicarbonate is also reabsorbed, presumably in the straight proximal tubule. With in vitro perfusion, this segment, but not the thick ascending limb of Henle, can reabsorb bicarbonate (11, 15).

Approximately 90% or more of filtered bicarbonate has been reabsorbed prior to reaching the early distal convoluted tubule, and a substantial portion of the remainder is reclaimed at this site. Furthermore heightened bicarbonate delivery results in increased reabsorption by the distal convoluted tubule, which indicates that its reabsorptive capacity is unsaturated under normal conditions (1). Carbonic anhydrase is absent from the luminal border in the distal nephron; and it has been suggested that a disequilibrium pH occurs at this site (19, 20). However, for the reason noted previously this requires revalidation. The pH of tubular fluid in this segment can be in the low 6.0 range, therefore some titratable acid and ammonium can accumulate.

Nevertheless, the ultimate site for the determination of net acid excretion is the collecting system.[2] In this portion of the nephron the remaining bicarbonate is reabsorbed, the steep pH gradients found in the final urine are generated, and titratable acid and ammonium accumulate in excess of bicarbonate. These actions account for the net acid excreted in the final urine. Recently it was shown that cortical collecting tubules taken from rabbits ingesting alkali are capable of secreting bicarbonate into the tubular fluid (13), and that turtle bladders (which resemble the mammalian collecting duct) also have a bicarbonate secretory mechanism (2, 22). Thus, the collecting system may serve as the final determinant for renal alkali as well as acid excretion.

Anuran Membranes

Potential insight into the mechanisms governing hydrogen ion transport by the mammalian collecting duct has been provided by studies of anuran membranes (2, 4, 5). These membranes resemble collecting duct epithelia phylogenetically and anatomically and also appear to be functionally similar (2, 23).

---

[2] Collecting system is used to refer to the cortical collecting tubules and the cortical and medullary collecting ducts.

Both turtle and toad bladders actively transport hydrogen ion, and the rate of flux depends on the electrochemical gradient against which secretion proceeds. The maximal chemical gradient achievable is approximately 3.0 pH units, i.e. a mucosal pH of 4.4 with a serosal pH of 7.4. With a gradient of this magnitude the hydrogen ion secretory flux rate is virtually nil, and there is little backflux of hydrogen ion from mucosa to serosa. Therefore under normal conditions backflux of hydrogen ion does not play a significant role in modifying net hydrogen ion secretion. If the mucosal pH is decreased to less than 4.4 by the addition of hydrogen ion, backflux increases; and it can occur at higher pH values when the membrane is pathologically altered, as for example with amphotericin B (2). Elevation of the mucosal fluid pH above 4.4 results in a progressive increase in the rate of hydrogen ion secretion: Clearly hydrogen ion secretion depends on the transmural pH, or chemical, gradient. Similar pH dependence of hydrogen ion secretion has been demonstrated in proximal and distal convoluted tubules of the kidney, but the collecting system has not been investigated (24).

The hydrogen ion secretory process is also dependent on the transepithelial electrical gradient (25, 26). Imposing a large mucosal positive gradient can completely obliterate hydrogen ion secretion, while the rate of transport is stimulated by removing this opposing force and actually enhanced by the presence of a negative potential (25, 26). Sodium transport is not directly coupled to hydrogen ion secretion in these epithelia, but it appears to be indirectly coupled by virtue of the transepithelial voltage (2, 5, 25, 26).

Of interest is the fact that aldosterone alters hydrogen ion transport by anuran membranes independent of its effects on sodium (27, 28). It increases the hydrogen ion secretory rate at any given electrochemical gradient, but does not alter the maximal gradient against which hydrogen ion flux can proceed. Thus aldosterone alters the capacity (capacitance) of the hydrogen pump, but does not alter the maximal force (proton motive force).

MECHANISMS FOR THE RENAL TUBULAR REGULATION OF ACID EXCRETION

Assuming that these features of anuran urinary epithelia are applicable to the mammalian collecting system, it is possible to envision several ways in which acid excretion could be modified.

Effect of Bicarbonate Delivery to the Distal Nephron on Net Acid Excretion

If the pump has a finite capacity for hydrogen ion secretion, the rate of bicarbonate delivery to the collecting system could dictate acid excretion. At high rates of bicarbonate delivery, hydrogen secretion will be utilized to reabsorb bicarbonate and net acid excretion will be minimized or ablated. Conversely, reduced bicarbonate delivery to the distal nephron resulting from a decrease in filtered load or an increase in proximal reabsorption could increase acid excretion. Micropuncture data indicate that bicarbonate concentration is decreased at the last accessible portion of the proximal convoluted tubule and in the distal convoluted tubule of acidotic rats (1). Thus a greater fraction of hydrogen ion secreted at distal sites can be utilized for net acid formation. This mechanism might facilitate increased net acid excretion, but its importance would depend upon whether the hydrogen ion secretory capacity of the collecting system is actually the rate-limiting event for net acid excretion with normal amounts of bicarbonate delivery. This crucial issue is currently unresolved. Nevertheless, with $NaHCO_3$ loading it seems likely that the bicarbonate delivery rates overcome the capacity of the distal nephron to secrete hydrogen ion, and account, at least in part, for the alkali diuresis.

Reabsorption of bicarbonate by the proximal tubule is enhanced by extracellular fluid volume contraction, potassium depletion, and an increase in pCO_2; it is diminished by volume expansion, hypocapnia, and parathyroid hormone (8). If reduced delivery of bicarbonate to the distal nephron increases net acid excretion, this could explain the changes in net acid excretion seen under certain conditions with increased proximal bicarbonate reabsorption, as for example hypercapnia.

When pCO_2 is elevated net acid excretion increases initially while plasma bicarbonate levels are rising, and returns to control rates after several days when a stable elevation of plasma bicarbonate is achieved (29). Hypercapnia increases bicarbonate reabsorption proximally, but its effects on the distal convoluted tubule are not clear (1, 20, 30, 31). Thus at the onset of hypercapnia enhanced reabsorption at proximal sites may decrease distal delivery and increase net acid excretion. Subsequently when plasma bicarbonate concentration and the filtered load are elevated, distal bicarbonate delivery may return to normal ablating the increase in net acid excretion observed in the early phase. However, it should be reemphasized that this sequence of events is predicated on an interplay between bicarbonate delivery to the collecting system and net acid excretion, which is currently unproven.

Intrinsic Modification in Hydrogen Ion Pump

Acid excretion could be modified by changing the intrinsic capacity of the hydrogen ion pump. This could occur by altering either the capacitance or proton motive force[3] of the pump; however, only the former change has been observed experimentally with isolated epithelia (2, 5). Furthermore a change in proton motive force seems unlikely, since there is no evidence to suggest that the maximal urine to blood pH gradient is enhanced in acidotic animals.

Whether or not acidosis modifies the capacity of the pump in the collecting duct is unclear. Anuran membranes from acidotic animals appear to have an increased capacity to secrete hydrogen ion (32, 33). Isolated collecting tubules from rabbits pretreated with NH_4Cl absorb bicarbonate, while those pretreated with $NaHCO_3$ secrete bicarbonate (13). Thus, in the rabbit there is some adaptive phenomenon in a hydrogen ion and/or bicarbonate secretory process. The implications in reference to the regulation of net acid excretion in other species are undefined. It is of interest, however, that the adaptive process appears to be exactly the opposite of the change found in the rat proximal convoluted tubule, where chronic acidosis suppresses net bicarbonate reabsorption (34).

As noted earlier aldosterone increases the capacitance of the hydrogen ion pump in anuran membranes via a direct sodium-independent process. In addition, studies in man and intact animals suggest that aldosterone secretion increases during acidosis (35, 36). Thus acidosis might modify hydrogen ion secretion, via an effect on the hydrogen ion pump mediated by an increase in aldosterone. Whether or not this mechanism plays a role in acid-base control has not been delineated.

Buffer Availability: Modification of the pH (Chemical) Gradient

The availability of nonbicarbonate buffer could influence the rate of hydrogen ion excretion by modifying the pH gradient against which hydrogen ion is secreted. An increase in phosphate excretion, for example, can increase titratable excretion (37); however, there is no evidence that a persistent increase in phosphate excretion is an important mechanism for dealing with elimination of an acid load.[4]

On the other hand, considerable evidence indicates that renal ammonia production is enhanced in acidosis (38). The increase in ammoniagenesis with chronic acidosis is well known, although the key biochemical events mediating this change are still in dispute (38). Recent evidence from intact rats and dogs, and isolated perfused rat kidneys, indicates that ammonia production also increases promptly in response to both acute metabolic and respiratory acidosis (39–41). Thus a decrease in systemic pH, in some fashion acutely increases ammoniagenesis. In view of the dynamics of hydrogen ion secretion in anuran membranes, increased availability of ammonia to buffer secreted hydrogen ions could promote increased net acid excretion.

---

[3] An increase in proton motive force results not only in the ability to achieve a steeper electrochemical gradient, but also in increased hydrogen ion secretion at any given gradient.

[4] One exception might be the ingestion of an acid phosphate load.

Alterations in the Electrical Gradient: Anions and Sodium Reabsorption

Finally, acid excretion could be modified by influencing the transepithelial voltage gradient in the collecting duct.

Since a fixed hydrogen ion is ingested with an anion, which ultimately is destined for excretion, a hydrogen load is associated with increased anion delivery to the collecting system. If the anion is poorly reabsorbable, it may play a role in potentiating hydrogen ion secretion by making the electrical gradient more favorable (42). In addition, acidosis decreases proximal sodium reabsorption, resulting in increased sodium delivery to the distal nephron (34). The impact of increased sodium reabsorption on hydrogen ion secretion by the collecting duct has not been clearly delineated, but experiments with anuran membranes raise the possibility that hydrogen ion secretion might be enhanced due to an altered electrical gradient. Observations in the intact animal suggest that distal sodium delivery influences acid excretion (43).

An increase in pCO_2 might also facilitate acid excretion by modifying the electrical gradient. It has been speculated that the primary effect of hypercapnia is to diminish chloride reabsorption (43). If so, increased chloride delivery to the collecting duct could act in the same fashion as impermeant anions to stimulate hydrogen ion secretion.

SIGNAL-INITIATING MODIFICATION OF RENAL ACID EXCRETION

It is apparent from the foregoing discussion that the study of anuran membranes, and the application of micromethodology to intact kidneys and tubular segments, has done much to further our knowledge of hydrogen ion transport mechanisms and the role of different nephron segments. Although these data permit identification of potential control mechanisms, they do not provide information concerning what specifically signals the kidney to alter renal acid excretion on a day-to-day basis, or how this message is translated into excretion of precisely the appropriate quantity of acid or alkali to maintain acid-base homeostasis. Some answers to these questions are suggested by long-term experiments in conscious animals; however, the construction of an overall scheme for the regulation of acid excretion must still be speculative.

It is generally tacitly assumed that systemic pH is the proximate event signaling the kidney to alter acid excretion, since acidosis increases and alkalosis diminishes acid excretion. In a recent provocative review Schwartz & Cohen (43) emphasized the substantial evidence suggesting that acid excretion by the kidney is not responsive to changes in systemic pH.

For example, variations in normal dietary intake can result in renal excretion of either alkali or acid, without detectable changes in systemic pH. The changes in pH could be so small that they escape detection; but this would require an exquisitely sensitive mechanism for their recognition, which to date has not been uncovered. It is widely appreciated that, at comparable degrees of acidosis, substantially greater quantities of acid are excreted with metabolic than with respiratory acidosis. Obviously systemic pH alone cannot dictate the difference, nor is intracellular pH a likely candidate, since respiratory disorders appear to cause greater deviations in intracellular pH than metabolic abnormalities (44, 45).

Finally, as demonstrated by Schwartz and co-workers the response to an acid load is highly dependent on the nature of the accompanying anion (46, 47). Acid excretion is comparable in response to sulfuric and to hydrochloric acid ingestion, but occurs at a higher plasma bicarbonate concentration and pH with the sulfate anion (46). The independence of acid excretion from systemic pH is demonstrated even more strikingly when animals with metabolic alkalosis induced by chloride depletion ingest nitric acid. In this setting the acid load is excreted quantitatively despite the persistence of an alkaline systemic pH (47). These observations all suggest that acid excretion is regulated by factors other than systemic pH. It should be emphasized, however, that these findings cannot necessarily be interpreted as indicating that systemic pH has no influence whatsoever on acid excretion. I believe it may within the constraints of other regulatory factors as is discussed subsequently.

SCHEME FOR REGULATION OF ACID EXCRETION

If systemic pH does not exert sole control of acid excretion, what may be the underlying mechanism(s)? Crucial insight into this question is gained by an examination of the studies in both normal and alkalotic animals subjected to acid loads of varying anion composition. As indicated above, in the normal animal hydrogen ion accompanying sulfate is excreted more efficiently than hydrogen ion accompanying chloride; i.e. acid excretion is comparable in steady-state conditions with both acid loads, but a higher plasma bicarbonate concentration and pH are maintained with sulfate (46). Similarly nitric acid is quantitatively excreted when administered to an alkalotic dog, but hydrochloric acid given to a similar animal results in sustained correction of the metabolic alkalosis (47). In both circumstances the different hydrogen excretory response can be explained based on the renal handling of the anions administered. Sulfate and nitrate are poorly reabsorbed by the renal tubule, and rapidly excreted in the urine. By contrast chloride is partially retained by the normal dog and completely retained by the alkalotic animal until the alkalosis is corrected. These observations suggest that excretion of anion into the urine is one mechanism for regulating acid excretion.

It is important to recognize, however, that anion leak alone does not dictate hydrogen ion excretion. If sulfate is substituted for chloride in the diet of a normal animal no significant change in net acid excretion or hydrogen ion homeostasis is observed (48). Rather the amount of anion in the urine sets a quantitative upper limit on the amount of acid that can be excreted. Within this limit, other mechanisms must exist to determine which cations, i.e. sodium, potassium, or hydrogen ion, will accompany the excreted anion. It could be pointed out that this is simply an obvious statement of the constraints of electroneutrality. However, by delineating an upper limit for hydrogen ion excretion, this provides a useful regulatory device, which reduces the need for an elegant hydrogen-ion-sensing mechanism. For example, the hydrogen ion in the daily diet is accompanied by anions that readily leak into the urine such as sulfate, phosphate, and organic acids. Thus recognition of minor changes in systemic hydrogen ion status may be unnecessary in order to deal with day-to-day fluctuations in acid intake.

Given this premise concerning the influence of anions on hydrogen ion secretion, the manner in which cation excretion is partitioned must be confronted. What mechanisms relegate a portion of the cation package to sodium, potassium, or hydrogen ion?

Although many questions related to the regulation of sodium excretion are still unanswered, it seems clear that the status of extracellular fluid (ECF) volume feeds back to appropriately regulate renal sodium excretion (49). Sodium handling is probably modulated at all sites along the nephron from the proximal tubule to collecting duct. The summation of these processes is the maintenance of sodium homeostasis such that chronic ingestion of Na_2SO_4 results largely in Na_2SO_4 excretion (48), whereas ingestion of H_2SO_4 results largely in H_2SO_4 excretion (46).

It has been suggested that when filtered sodium is reclaimed at sites in the distal rather than proximal nephron that hydrogen ion excretion can be promoted (43, 50–52). Expansion of ECF volume by vasopressin administration leads to increased acid excretion and correction of the acidosis in animals ingesting HCl (49, 50, 52), and this effect is abolished in adrenalectomized animals (51). In this experimental setting the hypotonicity appears to override volume expansion and stimulates aldosterone secretion. The volume-expansion-mediated increase in sodium delivery to the distal nephron, which can avidly reabsorb sodium, is thought to account for the heightened acid excretion. However, the increase in distal sodium delivery is accompanied by an increase in distal chloride delivery; and an increase in anion excretion, in the form of chloride, accompanies the increase in acid excretion. Thus the interplay between anion excretion and sodium reabsorption, appears to be an integral part of this phenomenon, and is one of the critical determinants of acid excretion. In this unique experimental setting, tonicity as well as volume appears to modulate sodium excretion. However, under most circumstances volume appears to predominate in determining whether or not sodium will be lost when anion excretion is promoted.

As depicted in Figure 1, if it is assumed that the portion of anion excretion accompanied by sodium is dictated by the intricacies of the volume–sodium homeostasis mechanism, the remainder must be excreted with potassium and hydrogen ion. The final pertinent issue is, "What governs the amount of hydrogen ion and potassium excreted within the defined constraints of the total anion excretion unaccompanied by sodium?" The experimental data available suggest that renal ammonia production may provide one important mechanism for regulating this partition between hydrogen ion and potassium.

As noted earlier, ammonia production is regulated by hydrogen ion homeostasis (38). Recent experiments in intact animals and isolated per-

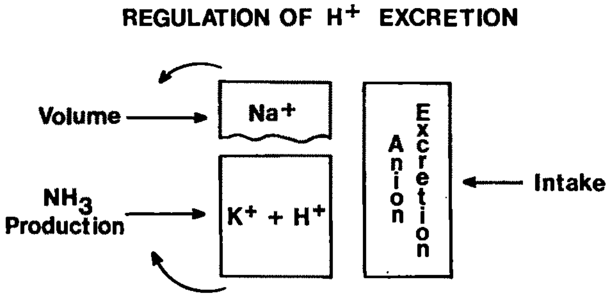


Figure 1 Scheme for the overall regulation of acid excretion. Anion excretion sets the quantitative upper limit for total cation excretion. Within this constraint, sodium excretion is controlled by volume-mediated regulatory mechanisms, and the partition between potassium and hydrogen ion excretion is modulated by renal ammonia production.

fused rat kidneys indicate that acute metabolic and respiratory acidosis stimulate and acute alkalosis depresses renal ammonia production. It is well recognized that chronic metabolic acidosis results in a sequence of renal metabolic adaptations that enhance renal ammonia production. However, the impact of chronic respiratory acidosis on renal ammonia production is not well defined (38).

Furthermore, an increase in renal ammonia production stimulates acid excretion (53). The mechanism is presumably via buffering of hydrogen in the urine, maintenance of a more favorable gradient for hydrogen ion secretion, and therefore persistance of a higher hydrogen ion flux rate. Thus ammonia production provides one mechanism through which acid excretion can be altered in direct response to systemic pH.

However, ammonia production can be modified by potassium as well as by pH (54). Potassium depletion, apparently even of minimal degree, stimulates renal ammonia production (55–57); ingestion of a high potassium diet, or potassium retention induced in other fashions, appears to diminish renal ammonia production (57–62). In addition, it appears that when ammonia production alters ammonium excretion it also modifies potassium excretion. An acute increase in ammonia production decreases potassium excretion in both humans and rats, and the event appears to be mediated by the collecting duct (53, 63). Thus a potassium deficit increases ammonia production, which in turn diminishes potassium excretion; a potassium surfeit may act in the opposite fashion.

Therefore hydrogen ion and potassium homeostasis can both modulate ammonia production, which can determine, at least in part, the partition between hydrogen ion and potassium excretion. It is important to recognize that this ammonia-regulated interplay takes place within the constraints of anion excretion unaccompanied by sodium. Thus hyperkalemia may diminish ammonia production but not decrease net acid excretion, if the quantity of anion excreted without sodium is increased. However, in this case a greater fraction of cation excretion is accounted for by potassium than hydrogen ion, consistent with the hypothesis that inhibition of ammonia production by hyperkalemia can facilitate potassium excretion (58). Similarly, potassium depletion can increase ammonia production but not increase acid excretion if sodium-free anion excretion is decreased (64). In this case the fraction of cation excretion accounted for by potassium is diminished, consistent with a potassium-sparing effect of enhanced ammonia formation.

It should be noted that the ammonia production system would not be expected to play a role in hydrogen ion or potassium regulation when the distal nephron is confronted with an alkali surfeit. Under this circumstance, the options available are excretion of the excess bicarbonate with either sodium or potassium, or its reclamation. Thus the interplay between potassium and acid-base regulation by the distal nephron is dictated by bicarbonate reabsorption rather than by acid excretion, and the role of ammonia production is inconsequential.

Finally, it must be stressed that this potential mechanism for altering potassium and acid excretion does not eliminate the possibility that other modulators of potassium and/or hydrogen ion excretion exist. It is quite conceivable that pH directly alters the capacitance of the hydrogen ion pump, as has recently been demonstrated in anuran membranes (65), that potassium homeostasis influences a potassium secretory pump (66), and/or that other coupling phenomena such as hydrogen ion mediation of potassium entry into renal tubular cells are important (66). The existence of these, or other alternative mechanisms of potassium and hydrogen ion control, would not mitigate against a simultaneous role for ammonia production in the regulatory process. One future challenge is to discern the relative importance of these potential modulators.

CONCLUSION

Enhanced comprehension of the mechanisms that regulate acid excretion has been obtained by observations in intact animals and humans, by the study of selective segments of the mammalian nephron in vivo and in vitro, and by the delineation of hydrogen ion transport characteristics of anuran membranes that resemble the epithelium of the mammalian collecting system. Acid excretion is ultimately determined by the hydrogen ion secretory mechanisms of the collecting system. It may be modified by bicarbonate handling at more proximal sites in the nephron, by actual modifications in the hydrogen ion secretory pump, by alterations in transepithelial voltage secondary to modifications in sodium transport and the characteristics of the anions in tubular fluid, or by alterations in the hydrogen ion buffering capacity mediated by changes in renal ammonia production. Our current knowledge has sharpened the focus on these potential regulatory mechanisms, but has not clearly resolved the importance of each component.

A speculative scheme for the regulation of acid excretion is proposed in which urinary anion excretion defines an upper limit for cation excretion. Within this constraint, volume-related mechanisms dictate the amount of sodium excreted, and ammonia production, which is regulated by both pH and potassium homeostasis, modulates the relationship between acid and potassium excretion. Like the specifics of hydrogen ion transport at a renal tubular level, this hypothesis for overall regulation of acid excretion is a functional model subject to further innovative experimental challenges.

ACKNOWLEDGMENTS

The author would like to thank Ms. Gloria Agas and Mrs. Laura Larisey for their help in preparation of the manuscript. These studies were supported in part by PHS Grant AM 25248 and Grant 77–971 from the American Heart Association.

Literature Cited

1. Malnic, G., Giebisch, G. 1972. Mechanism of renal hydrogen ion secretion. *Kidney Int.* 1:280–96
2. Steinmetz, P. R. 1974. Cellular mechanisms of urinary acidification. *Physiol. Rev.* 54:890–956
3. Rector, F. C. Jr. 1976. Renal acidification and ammonia production; chemistry of weak acids and bases; buffer mechanisms. In *The Kidney,* ed. B. M. Brenner, F. C. Rector, Jr., pp. 318–43. Philadelphia, London & Toronto: Saunders. 1941 pp.
4. Malnic, G., Steinmetz, P. R. 1976. Transport processes in urinary acidification. *Kidney Int.* 9:172–88
5. Al-Awqati, Q. 1978. H^+ transport in urinary epithelia. *Am. J. Physiol.* 235:F77–88
6. O'Connor, G., Kunau, R. T. 1978. Renal transport of hydrogen and potassium. *Contemp. Iss. Nephrol. Acid-Base and Potassium Homeostasis* 2:1–29
7. Gennari, F. J., Cohen, J. J. 1978. Renal tubular acidosis. *Ann. Rev. Med.* 29:521–41
8. Levine, D. Z. 1978. Difficulties in the micropuncture evaluation of proximal renal bicarbonate reabsorption: an overview for the general reader. *Can. J. Physiol. Pharmacol.* 56:354–61
9. Warnock, D. G., Burg, M. B. 1977. Urinary acidification: CO_2 transport by the rabbit proximal straight tubule. *Am. J. Physiol.* 232:F20–25
10. Burg, M., Green, N. 1977. Bicarbonate transport by isolated perfused rabbit proximal convoluted tubules. *Am. J. Physiol.* 233:F307–14
11. McKinney, T. D., Burg, M. B. 1977. Bicarbonate and fluid absorption by renal proximal straight tubules. *Kidney Int.* 12:1–8
12. McKinney, T. D., Burg, M. B. 1978. Bicarbonate absorption by rabbit cortical collecting tubules in vitro. *Am. J. Physiol.* 234:F141–45
13. McKinney, T. D., Burg, M. B. 1977. Bicarbonate transport by rabbit cortical collecting tubules. Effect of acid and alkali loads in vivo on transport in vitro. *J. Clin. Invest.* 60:766–68
14. McKinney, T. D., Burg, M. B. 1978. Bicarbonate secretion by rabbit cortical collecting tubules in vitro. *J. Clin. Invest.* 61:1421–27
15. Iino, Y., Burg, M. B. 1979. Effect of in vivo acidosis and alkalosis on in vitro bicarbonate transport by rabbit proximal straight tubules and cortical thick ascending limbs. *Clin. Res.* 27:418A
16. Jacobson, H. R., Seldin, D. W., Kokko, J. P. 1979. Comparison of early superficial and juxtamedullary proximal convolutions. *Clin. Res.* 27:498A
17. Lucci, M. S., Warnock, D. G., Rector, F. C. Jr. 1979. Carbonic anhydrase-dependent bicarbonate reabsorption in the rat proximal tubule. *Am. J. Physiol.* 236:F58–65
18. Lucci, M. S., DuBose, T. D. Jr., Pucacco, L. R., Kokko, J. P., Carter, N. W. 1979. Effect of carbonic anhydrase inhibitors of varying permeabilities on proximal tubular pH and total CO_2 reabsorptive rate. *Clin. Res.* 27:423A
19. Rector, F. C. Jr., Carter, N. W., Seldin, D. W. 1965. The mechanism of bicarbonate reabsorption in the proximal and distal tubules of the kidney. *J. Clin. Invest.* 44:278–90
20. Vieira, F. L., Malnic, G. 1968. Hydrogen ion secretion by rat renal cortical tubules as studied by an antimony microelectrode. *Am. J. Physiol.* 214:710–18
21. DuBose, T. D. Jr., Pucacco, L. R., Seldin, D. W., Carter, N. W., Kokko, J. P. 1978. Direct determination of pCO_2 in the rat renal cortex. *J. Clin. Invest.* 62:338–48
22. Leslie, B. R., Schwartz, J. H., Steinmetz, P. R. 1973. Coupling between Cl^- absorption and HCO_3 secretion in turtle urinary bladder. *Am. J. Physiol.* 225:610–17
23. Terao, N., Tannen, R. L. 1978. Urinary acidification by the isolated perfused rat kidney. *Proc. Am. Soc. Nephrol.* 11:149A

24. Giebisch, G., Malnic, G., de Mello, G. B., de Mello Aires, M. 1977. Kinetics of luminal acidification in cortical tubules of the rat kidney. *J. Physiol.* 267:571–99
25. Al-Awqati, Q., Mueller, A., Steinmetz, P. R. 1977. Transport of H^+ against electrochemical gradients in turtle urinary bladder. *Am. J. Physiol.* 233: F502–8
26. Ziegler, T. W., Fanestil, D. D., Ludens, J. H. 1976. Influence of transepithelial potential difference on acidification in the toad urinary bladder. *Kidney Int.* 10:279–86
27. Ludens, J. H., Fanestil, D. D. 1974. Aldosterone stimulation of acidification of urine by isolated urinary bladder of the Colombian toad. *Am. J. Physiol.* 226:1321–26
28. Al-Awqati, Q., Norby, L. H., Mueller, A., Steinmetz, P. R. 1976. Characteristics of stimulation of H^+ transport by aldosterone in turtle urinary bladder. *J. Clin. Invest.* 58:351–58
29. Polak, A., Haynie, G. D., Hays, R. M., Schwartz, W. B. 1961. Effects of chronic hypercapnia on electrolyte and acid-base equilibrium. I. Adaptation. *J. Clin. Invest.* 40:1223–39
30. de Mello Aires, M., Malnic, G. 1976. Peritubular pH and pCO_2 in renal tubular acidification. *Am. J. Physiol.* 228: 1766–74
31. Levine, D. Z. 1971. Effect of acute hypercapnia on proximal tubular water and bicarbonate reabsorption. *Am. J. Physiol.* 221:1164–70
32. Frazier, L. W., Vanatta, J. C. 1971. Excretion of H^+ and NH_4^+ by the urinary bladder of the acidotic toad and the effect of short-circuit current on the excretion. *Biochem. Biophys. Acta* 241: 20–29
33. Ziegler, T. W., Ludens, J. H., Fanestil, D. D. 1974. Role of carbonic anhydrase in urinary acidification by the toad bladder. *Am. J. Physiol.* 227:1132–38
34. Levine, D. Z., Nash, L. A. 1973. Effect of chronic NH_4Cl acidosis on proximal tubular H_2O and HCO_3 reabsorption. *Am. J. Physiol.* 225:380–84
35. Perez, G. O., Oster, J. R., Vaamonde, C. A., Katz, F. H. 1977. Effect of NH_4Cl on plasma aldosterone, cortisol and renin activity in supine man. *J. Clin. Endocrinol. Metab.* 45:762–67
36. Perez, G. O., Oster, J. R., Kem, D. C., Vaamonde, C. A. 1979. Effect of acute metabolic acidosis on the renin-aldosterone system: Mechanism of increases in plasma aldosterone. *Clin. Res.* 27:499A
37. Schiess, W. A., Ayer, J. L., Lotspeich, W. D., Pitts, R. F. 1948. The renal regulation of acid-base balance in man. II. Factors affecting the excretion of titratable acid by the normal human subject. *J. Clin. Invest.* 27:57–64
38. Tannen, R. L. 1978. Ammonia metabolism. *Am. J. Physiol.* 235:F265–77
39. Emmett, M., Rascoff, J., Relman, A. S., Narins, R. G. 1976. Effects of acute acid-base changes on in vivo rat renal NH_3. *Proc. Am. Soc. Nephrol.* 9:96
40. Fine, A., Bennett, F. I., Alleyne, G. A. O. 1978. Effects of acute acid-base alterations on glutamine metabolism and renal ammoniagenesis in the dog. *Clin. Sci. Mol. Med.* 54:503–8
41. Tannen, R. L., Ross, B. D. 1979. Ammoniagenesis by the isolated perfused kidney. The critical role of urinary acidification. *Clin. Sci.* 56:353–64
42. Clapp, J. R., Rector, F. C. Jr., Seldin, D. W. 1962. Effect of unreabsorbed anions on proximal and distal transtubular potentials in rats. *Am. J. Physiol.* 202:781–86
43. Schwartz, W. B., Cohen, J. J. 1978. The nature of the renal response to chronic disorders of acid-base equilibrium. *Am. J. Med.* 64:417–28
44. Adler, S., Roy, A., Relman, A. S. 1965. Intracellular acid-base regulation. II. The interaction between CO_2 tension and extracellular bicarbonate in the determination of muscle cell pH. *J. Clin. Invest.* 44:21–30
45. Struyvenberg, A., Morrison, R. B., Relman, A. S. 1968. Acid-base behavior of separated canine renal tubule cells. *Am. J. Physiol.* 214:1155–62
46. De Sousa, R. C., Harrington, J. T., Ricanati, E. S., Shelkrot, J. W., Schwartz, W. B. 1974. Renal regulation of acid-base equilibrium during chronic administration of mineral acid. *J. Clin. Invest.* 53:465–76
47. Tannen, R. L., Bleich, H. L., Schwartz, W. B. 1966. The renal response to acid loads in metabolic alkalosis: An assessment of the mechanisms regulating acid excretion. *J. Clin. Invest.* 45:562–72
48. Stinebaugh, B., Miller, R. B., Relman, A. S. 1969. The influence of non-reabsorbable anions on acid excretion. *Clin. Sci.* 36:53–65
49. Klahr, S., Slatopolsky, E. 1973. Renal regulation of sodium excretion. Function in health and in edema-forming states. *Arch. Intern. Med.* 131:78–91
50. Lowance, D. C., Garfinkel, H. G., Mattern, W. D., Schwartz, W. B. 1972. The effect of chronic hypotonic volume ex-

pansion on the renal regulation of acid-base equilibrium. *J. Clin. Invest.* 51:2928–40

51. Cohen, J. J., Hulter, H. N., Smithline, N., Melby, J. C., Schwartz, W. B. 1976. The critical role of the adrenal gland in the renal regulation of acid-base equilibrium during chronic hypotonic expansion. *J. Clin. Invest.* 58:1201–8
52. Smithline, N. 1977. Evidence that alterations in distal sodium delivery and distal sodium avidity affect the rate of excretion of acute HCl load. *Kidney Int.* 11:181–89
53. Tannen, R. L., Terrien, T. 1975. Potassium-sparing effect of enhanced renal ammonia production. *Am. J. Physiol.* 228:699–705
54. Tannen, R. L. 1977. Relationship of renal ammonia production and potassium homeostasis. *Kidney Int.* 11:453–65
55. Tannen, R. L. 1970. The effect of uncomplicated potassium depletion on urine acidification. *J. Clin. Invest.* 49:813–27
56. Baertl, J. M., Sancetta, S. M., Gabuzda, G. H. 1963. Relation of acute potassium depletion to renal ammonium metabolism in patients with cirrhosis. *J. Clin. Invest.* 42:696–706
57. Tannen, R. L., McGill, J. 1976. The influence of potassium on renal ammonia production. *Am. J. Physiol.* 231:1178–84
58. Tannen, R. L., Wedell, E., Moore, R. 1973. Renal adaptation to a high potassium intake: The role of hydrogen ion. *J. Clin. Invest.* 52:2089–2101
59. Kamm, D. E. 1971. Dissociation of urine pH and NH_3 excretion during KCl and NaCl loading. *Proc. Am. Soc. Nephrol.* 5:36
60. Hulter, H. N., Ilnicki, L. P., Harbottle, J. A., Sebastian, A. 1977. Impaired renal H^+ secretion and NH_3 production in mineralocorticoid-deficient glucocorticoid-replete dogs. *Am. J. Physiol.* 232:F136–46
61. Sebastian A., Schambelan, M., Lindenfeld, S., Morris, R. C. Jr. 1977. Amelioration of metabolic acidosis with fludrocortisone therapy in hyporeninemic hypoaldosteronism. *N. Engl. J. Med.* 297:576–83
62. Szylman, P., Better, O. S., Chaimowitz, C., Rosler, A. 1976. Role of hyperkalemia in the metabolic acidosis of isolated hypoaldosteronism. *N. Engl. J. Med.* 294:361–65
63. Jaeger, P., Karlmark, B., Giebisch, G. 1978. Micropuncture study of the effects of NH_4 on tubular acidification and potassium transport. *Clin. Res.* 26:542A
64. Burnell, J. M., Teubner, E. J., Simpson, D. P. 1974. Metabolic acidosis accompanying potassium deprivation. *Am. J. Physiol.* 227:329–33
65. Cohen, L. H., Steinmetz, P. R. 1974. Cell [OH] as a determinant of H^+ secretion by turtle urinary bladder: asymmetry in the control of the H^+ pump. *Clin. Res.* 27:520A
66. Wright, F. S., Giebisch, G. 1978. Renal potassium transport: contributions of individual nephron segments and populations. *Am. J. Physiol.* 235:F515–27

Ann. Rev. Med. 1980. 31:51–57
Copyright © 1980 by Annual Reviews Inc. All rights reserved

AMANITA MUSHROOM POISONING

❖7342

D. H. Mitchel, M.D., F.A.C.P.

Department of Medicine, University of Colorado Medical Center,
Denver, Colorado 80210

INTRODUCTION

For a century following the extraction of muscarine from *Amanita muscaria* (1) all mushroom poisoning (mycetism) was attributed to muscarine, for which the antidote was atropine. This concept was wrong in two aspects: (*a*) Muscarine is not the main toxin that produces the hallucinogenic or soporific effects of *A. muscaria* poisoning; (*b*) While atropine is a specific antidote for muscarine poisoning, it does not relieve and may actually aggravate the symptoms of other types of mushroom poisoning. In the past ten years, many toxic substances have been extracted from mushrooms and it is now apparent that mushroom poisoning is a complicated problem, involving many chemically unrelated toxins that affect different organ systems of the body and require different specific treatments (see Table 1). The most lethal of these toxins are the amatoxins of the *Amanita phalloides* group of mushrooms. Perhaps because of the "muscarine myth" or perhaps because of the simplistic wish of the toxicologists for an antidote for each toxin, the term "Amanita poisoning" is now being used to characterize the more complicated problem of mushroom poisoning, and the most recent antidote being suggested is thioctic acid. Actually, the treatment of mushroom poisoning is no more mysterious or complicated than the treatment of any other poisoning and, unfortunately, the number of specific antidotes is no greater. The mainstay of treatment is that of supportive care as dictated by the condition of the patient (see Table 2).

0066-4219/80/0401-0051$01.00

Table 1 Common types of mushroom poisoning[a]

Type	Group	Toxins	Principle mushrooms	Onset of symptoms	Symptoms	Treatment and duration of illness
A	Toxins causing cellular destruction—delayed onset					
	I	Cyclopeptides (amanitins)	*Amanita bisporigera* *A. ocreata* *A. phalloides* *A. verna* *A. virosa* *Galerina autumnalis* *G. marginata* *G. venenata* *Lepiota* sp.	6–24 hours (typically 10–14 hr)	Abdominal pains, nausea, vomiting and diarrhea lasting 1+ days; short remission of symptoms; then recurrence of pains with jaundice, renal shutdown, convulsions, coma, and death	Thioctic acid 50–150 mg q 6 hr IV with glucose. Penicillin-G 250 mg/kg/day in continuous IV infusion. Vitamin K 40 mg IV daily. Vitamin B complex IV. Massive doses of corticosteroids, e.g. Dexamethasone 20–40 mg IV daily. Maintain fluid and electrolyte balance. Follow liver and renal parameters and blood sugar. Death may occur on 4th to 7th days or recovery may take two weeks.
	II	Gyromitrin, Monomethylhydrazine (MMH)	*Gyromitra esculenta* and others	6–12 hours	Bloated feeling, nausea, vomiting, watery (or bloody) diarrhea, abdominal pains, muscle cramps, faintness, loss of coordination and, in severe cases, convulsions, coma, and death	Pyridoxine HCl 25 mg/kg IV titrated with patient's symptoms. Follow methemoglobin and free hemoglobin levels and hepatic parameters. Death may occur on the 5th to 7th days.
B	Toxins affecting the autonomic nervous system—rapid onset					
	III	Muscarine and other muscarinic compounds	*Clitocybe dealbata*, *C. cerussata*, and perhaps *C. illudens* (*Omphalotus olearius*) and others; *Inocybe* most sp.	0.5–2 hours	PSL syndrome (Perspiration, salivation, and lacrimation), blurred vision, abdominal cramps, watery diarrhea, constriction of pupils, fall in blood pressure, slow pulse	Atropine 1–2 mg IV repeated p.r.n. Symptoms subside within 6–24 hours. Death reported only in people with concurrent disease.

	IV	Coprine (Antabuse-like)	*Coprinus atramentarius*, *Clitocybe clavipes*	About 30 minutes after drinking alcohol; as long as 5 days after eating mushrooms.	Flushing of face and neck, distention of neck veins, swelling and tingling of hands, metallic taste, tachycardia and hypotension; later, nausea, vomiting, and sweating	Avoid elixirs and tinctures. Propranolol may be necessary to control arrhythmias. Recovery usually spontaneous within 2–4 hours.
C	Toxins affecting the central nervous system—rapid onset					
	V	Ibotenic acid, Muscimol (isoxazoles)	*Amanita muscaria*, *A. pantherina*, and others	0.5–2 hours	Dizziness, incoordination, staggering (intoxication); muscular jerking and spasms; hyperkinetic activity; coma-like deep sleep and "visions"	Do *not* give atropine unless definate cholinergic symptoms are present. Do *not* intubate unless there is actual respiratory distress. Recovery in 4–24 hours.
	VI	Psilocybin and Psilocin (indoles)	*Psilocybe cubensis*, *P. baeocystis*, and many others	30–60 (180) minutes	Pleasant or apprehensive mood; unmotivated laughter; hilarity; compulsive movements; muscle weakness; drowsiness; "visions" while awake, then sleep	Diazepam rarely necessary. Reassurance for apprehension. Recovery within 6 hours.
D	Toxins affecting the gastrointestinal tract—rapid onset					
	VII	Diverse, mostly unknown	Many species from diverse genera	0.5–2 (4) hours	Principally nausea, vomiting, diarrhea, and abdominal pain	Careful attention to fluid and electrolyte replacement. Death rare; recovery time varies according to species from an hour to several days, to a week.

---

[a] Modified from *Toxic and Hallucinogenic Mushroom Poisoning: A Handbook for Physicians and Mushroom Hunters* by G. H. Lincoff and D. H. Mitchel © 1977 by Litton Educational Publishing, Inc. Reprinted by permission of Van Nostrand Reinhold Co.

Table 2 Treatment of mushroom poisoning—general measures

I. Reduction of absorption
  A. Emesis, if not already adequate from poisoning itself. Ipecac syrup 15–30 cc followed by ±500 cc oral liquids
  B. Gastric lavage if patient is unconscious or convulsing. Large bore gastric tube (34 French or larger) with ±20 L NS
  C. Activated charcoal, 30 g in 60 cc water and repeat
  D. Catharsis, if not already adequate; magnesium sulfate 250 mg/kg p.o.
II. Enhancement of excretion
  A. Forced diuresis to obtain urine output of 3–6 cc/kg/hr with oral or IV fluids. Add furosemide 1 mg/kg/hr IV p.r.n.
  B. Hemodialysis with extracorporal circulation through charcoal filter if liver failure occurs, or hemodialysis if toxin molecule is small enough or if renal failure occurs
III. Supportive measures
  A. Maintain respiration with adequate airway, oxygen, and rarely, *in extremus*, ventilatory support
  B. Maintain blood pressure with IV fluids and, if necessary, vasopressors, e.g. Isoproterenol
  C. Maintain blood sugar and electrolytes
  D. Follow hepatic and renal parameters for at least 48 hours
  E. Antispasmodics and/or antiemetics after GI tract is empty
  F. Analgesics and/or opiates for pain
  G. Sedatives for anxiety, hysteria, hallucinations, or convulsions, which may occasionally occur in children, e.g. Diazepam
IV. Specific therapy with drug intervention is rarely required, and then only when the type of poisoning is diagnosed!

AMANITA POISONING

Toxins and Pathophysiology

Mushrooms in the *Amanita phalloides* group cause over 50% of all cases of mushroom poisoning, 95% of the fatal cases, and justly deserve their common names: "Death Cap" and "Destroying Angel." The study of the responsible toxins has been carried out by scientists in Central Europe for the past century. Over the past 40 years, the Wielands (2–5) and, more recently, Faulstich (5, 6) in Wieland's laboratories in Heidelberg, have identified and established the structural formulas of the toxic cyclopeptides involved. At least seven heptapeptide phallotoxins and eight octapeptide amatoxins have been identified. The phallotoxins are probably not involved in cases of human poisoning, but five of the amatoxins are among the most lethal toxins known. They inhibit nuclear RNA polymerase B (or II) and thus interfere with both RNA and DNA transcription (7). This in turn causes cellular necrosis and prevents reproduction of cells, especially those that receive the initial contact with the poison and those with the most rapid rates of turnover, for example, those in the gastrointestinal mucosa, liver, and kidneys.

Symptoms and Course

The symptoms of Amanita poisoning do not appear until 6 to 24 (usually 10–12) hours after the ingestion of the mushroom. Presumably, this latent period is the time necessary for the absorption of the toxin into the cells of the GI tract and the inhibition of RNA transcription to produce symptoms. After this characteristic latent interval, severe abdominal pains with nausea, vomiting, and profuse "cholera-like" diarrhea occur and last for about 24 hours. If, with electrolyte and fluid (up to 9 liters per day) replacement, the patient survives this acute episode, a 12–24-hour remission usually occurs. If the correct diagnosis of mushroom poisoning is not suspected and liver enzyme studies are not obtained, the patient may be discharged at this time under the mistaken assumption that he has recovered from an acute gastroenteritis of unknown etiology. Within 24 hours, rapidly increasing jaundice and renal shutdown may occur with concomitant elevation of liver enzymes, BUN, and creatinine. Severe or inadequately treated cases progress into coma, with its associated neurologic changes, and death occurs from the fourth to the seventh day. Recovery in serious cases may take as long as two to three weeks. Pathologic changes include necrosis of the intestinal mucosa, hepatic necrosis similar to "acute yellow atrophy," and renal necrosis particularly of the cells of the proximal tubule.

Therapy

The search for an antidote for the Amanita poisons has been as avid, if not as productive, as the study of the toxins. The story, too long to relate here, is reviewed in recent books (8, 9). Most of the therapies advised, from emulsions of rabbits' brains to horse serum antitoxins, have not stood the test of time. The only therapy universally accepted as beneficial is that of careful monitoring of the patient and replacement therapy as indicated, similar to that used for other types of hepatitis or acute tubular necrosis of the kidney. Steroids, antibiotics, intravenous vitamin B complex, and the like, which are of debatable value in the treatment of hepatitis and/or of tubular necrosis, are equally controversial in the treatment of Amanita poisoning. From these proposed therapies for hepatitis, the most recent recommended treatment for Amanita poisoning has evolved.

Reasoning that, if vitamins of the B complex aid the damaged liver, similar compounds involved in the Krebs cycle might also be beneficial, European physicians experimented with thioctic acid (alpha-lipoic acid), a coenzyme in the Krebs cycle. At an international symposium in Naples in 1955, it was reported to be of benefit in the treatment of liver disease in animals (10) and later it was used in humans with liver damage from heavy metal poisoning (11). Kubička (12) in 1968 reported from Czechoslovakia the use of thioctic acid in 40 cases of *Amanita phalloides* poisoning with only one death instead of the usual 50–95% mortality. Since its first use in the United States in 1970 (13), thioctic acid has been credited for the low mortality in many case reports of Amanita poisoning (14–17). More recently in both Europe and the United States, thioctic acid has been used and credited with some benefit in the treatment of other kinds of mushroom poisonings, e.g. *Gyromitra* sp. (18) and *Cortinarius* sp. (19). There have been no well-controlled studies, however, in humans or laboratory animals to prove or disprove the efficacy of thioctic acid therapy. Alleva (20), in experiments for the Food and Drug Administration (FDA), reported failure of thioctic acid to cure mushroom-poisoned mice and dogs, but Bartter (16, 21) criticized these experiments because hypoglycemia was not controlled in the test animals.

In 1976, Moroni et al (15) reported 100% survival of 33 individuals with Amanita poisoning and attributed the excellent results to the use of intravenous penicillin as well as thioctic acid and cortisone. The rationale for the use of penicillin was based on the experiments of Floersheim (22, 23), who showed that many substances, including penicillin, reduced the toxicity of amatoxins, theoretically by blocking their binding sites with proteins. Reasoning that protein-bound toxin would not filter through the glomerulus of the kidney, but would instead return to the systemic circulation to produce further liver damage, these Italian investigators elected to use intravenous penicillin rather than the other more toxic compounds Floersheim had tested. Faulstich (personal communication) and others (24) doubt this whole premise since they could not demonstrate any significant binding of the toxin to human serum albumin. Though Fiume (25) describes "linking of amanitin to bovine albumin" and "conjugation with albumin" in the sinusoids of the mouse liver and in the proximal convoluted tubule cells of the rat kidney, he states (24) that this does not occur in, or have any bearing on, human poisonings.

The FDA is now evaluating both thioctic acid and intravenous penicillin therapy in dog experiments in which replacement therapy is being carefully controlled. Until the results of these and/or other experiments show these drugs to be valueless, it is the opinion of this reviewer that they should be used as outlined in Table 1.

Literature Cited

1. Schmiedeberg, O., Koppe, R. 1869. *Das Muscarin.* Leipzig: Vogel. 111 pp.
2. Lynen, F., Wieland, U. 1937. Über die Giftstoffe des Knollenblätterpilzes. IV. Kristallisation von Phalloidin. *Justus Liebigs Ann. Chem.* 533:93–99
3. Wieland, H., Hallermayer, R. 1941. Über die Giftstoffe des Knollenblätterpilzes. VI. Amanitin, das Hauptgift des Knollenblätterpilzes. *Justus Liebigs Ann. Chem.* 548:1–7
4. Wieland, T., Wieland, O. 1972. The toxic peptides of Amanita species. In *Microbial Toxins: Vol. VIII Fungal Toxins,* ed. S. Kadis, A. Ciegler, S. J. Ajl, pp. 249–80. New York: Academic. 400 pp.
5. Wieland, T., Faulstich, H. 1978. Amatoxins, phallotoxins, phallolysin, and antamanide: the biologically active components of poisonous Amanita mushrooms. *CRC Crit. Rev. Biochem.* 5(3):185–260
6. Faulstich, H. 1977. Structure of poisonous components of *Amanita phalloides. Curr. Probl. Clin. Biochem.* 7:1–10
7. Lindell, T. J., Weinberg, F., Morris, P. W., Roeder, R. G., Rutter, W. J. 1970. Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. *Science* 170:447–49
8. Lincoff, G., Mitchel, D. H. 1977. *Toxic and Hallucinogenic Mushroom Poisoning.* New York: Van Nostrand Reinhold. 267 pp.
9. Rumack, B. H., Salzman, E., eds. 1978. *Mushroom Poisoning: Diagnosis and Treatment.* West Palm Beach, Florida: CRC Press. 263 pp.
10. International Symposium on Thioctic Acid, Naples, 1955: Thioctic Acid, Physics, Chemistry, and Biology. *Chem. Abst.* 51:8153–55 (1957)
11. Schmidt, U., Grafen, P., Goedde, H. W. 1965. Chemistry and biochemistry of alpha lipoic acid. *Angew. Chem. Int. Ed. Engl.* 4(10):845–56
12. Kubička, J. 1968. Traitement des empoisonnements fongiques phalloidiniens en Tchecoslovaquie. *Acta Mycol.* 4(2): 373–78
13. Finestone, A. J., Berman, R., Widmer, B., Markowitz, J., Laquer, U. J. 1972. Thioctic acid treatment of acute mushroom poisoning. *Pa. Med.* 75(7):49–51
14. Becker, C. E., Tong, T. G., Boerner, U., Roe, R. L., Scott, R. A. T., MacQuarrie, M. B., Bartter, F. 1976. Diagnosis and treatment of Amanita phalloides-type mushroom poisoning—Use of thioctic acid. *West. J. Med.* 125:100–9
15. Moroni, F., Fantozzi, R., Masini, E., Mannaioni, P. F. 1976. A trend in the therapy of *Amanita phalloides* poisoning. *Arch. Toxicol.* 36:111–15
16. Bartter, F. 1975. *Sci. Am.* 233(1):8
17. Teutsch, C., Brennan, R. W. 1978. Amanita mushroom poisoning with recovery from coma. *Ann. Neurol.* 3(2): 177–79
18. Mittman, W. 1968. Zur Klinik und Therapie der Lorchelvergiftung (*Gyromitra esculenta*). *Z. Aerztl. Fortbild.* 67:710
19. Berkson, B. 1978. *The Plain Dealer.* Cleveland, Ohio. Oct. 4, p. 8-A
20. Alleva, F. R., Balazs, T., Sager, A. O., Done, A. K. 1975. Failure of thioctic acid to cure mushroom-poisoned mice and dogs. *Toxicol. Appl. Pharmacol.* 33:122–99
21. Bartter, F. 1975. Thioctic acid and mushroom poisoning. *Science* 187:216
22. Floersheim, G. L. 1972. Antidotes to experimental α-amanitin poisoning. *Nature, New Biol.* 236:115–17
23. Floersheim, G. L. 1976. Antagonistic effects against single lethal doses of *Amanita phalloides. Naunyn-Schmiedeberg's Arch. Pharmacol.* 293:171–74
24. Fiume, L., Sperti, S., Montanaro, L., Busi, C., Costantino, D. 1977. Amanitins do not bind to serum albumin. *Lancet* 1:1111
25. Fiume, L. 1977. Mechanism of action of amanitins. *Curr. Probl. Clin. Biochem.* 7:20–28

Ann. Rev. Med. 1980. 31:59–77
Copyright © 1980 by Annual Reviews Inc. All rights reserved

THE MEDICAL TREATMENT OF GALLSTONES

◆7343

Ian A. D. Bouchier, M.D., F.R.C.P., F.R.C.P.(E.)

Department of Medicine, University of Dundee, Ninewells Hospital and Medical School, Dundee, Scotland

INTRODUCTION

Gallstones do not disappear spontaneously and medicine has long sought a safe and effective drug for their dissolution. After many excessive claims and false hopes there now is a dependable method for dissolving gallstones. The treatment is not applicable to all types of stone and there are many qualifications to its use, but it is a reasonable alternative to surgery for the treatment of cholelithiasis.

At present medical therapy is useful only for those gallstones comprised solely or mainly of cholesterol; these are the majority of stones occurring in the Western world. Drug treatment centers around chenodeoxycholic acid (CDCA). I shall review current concepts of how cholesterol is maintained in solution in bile, the pathogenesis of cholesterol gallstones, the experience with CDCA and other agents used to treat gallstones, and finally the nonsurgical management of common bile duct stones.

THE SOLUBILITY OF CHOLESTEROL IN BILE

Bile is 85–95% water. Cholesterol is insoluble in water and is maintained in solution in bile because of the formation of mixed micelles comprising cholesterol, phospholipids, and bile salts (1–4).

Using phase diagrams Admirand & Small (5) demonstrated that patients with cholesterol gallstones could be distinguished from non-gallstone individuals by the composition of their biliary lipids. The essential feature of bile-containing cholesterol stones is an excess of cholesterol relative to bile salts and phospholipids; such bile is unable to maintain the cholesterol in

0066-4219/80/0401-0059$01.00

a micellar solution. The fact that a number of studies have subsequently failed to separate cholesterol gallstone-forming bile from stone-free bile on the basis of bile physicochemistry (6–8) does not negate the fundamental precept that cholesterol supersaturation is necessary for lithogenesis.

Cholesterol supersaturation is essential to, but not the sole factor for, gallstone formation, the supersaturated bile itself originating from the liver. Whatever the influence of the gallbladder on lithogenesis, the basic defect in cholesterol gallstone disease is metabolic, existing in the liver (9–11). Cholesterol supersaturation can arise from an excess of biliary cholesterol or a relative lack of bile salts, or both (12, 13). In obese persons an increase in cholesterol synthesis and excretion appears to be the predominant event (14). But in the majority of gallstone-forming patients, diminished hepatic secretion rate of bile salts is the fundamental defect, and this can be related to a reduced total body pool of bile salts. A primary deficiency of hepatic biliary lecithin has never been established.

The reduction in bile salt pool size is unexplained. Neither altered steady-state synthesis nor absorption of bile salts is at fault, and currently the most acceptable view is that there is either an increased cycling frequency of bile salts within the enterohepatic circulation, thereby suppressing synthesis, or, alternatively, an oversensitive feedback inhibition of bile salt synthesis (15, 16).

Metabolic Defect in Cholesterol Gallstone Disease

Patients with cholesterol gallstones are believed to have a defect in cholesterol synthesis relative to bile salts, thereby producing an excess of biliary cholesterol. The hepatic activities of HMG-Co A reductase are higher and cholesterol 7α-hydroxylase lower in subjects with gallstones compared to those without (21–23). In addition, gallstone-forming persons have lower 12α-hydroxylase activity than normal. If this decrease were greater than that of cholesterol 7α-hydroxylase, it might account for the increased proportion of dihydroxy bile salts found in the bile of gallstone patients (24).

CHENODEOXYCHOLIC ACID

In 1971 Thistle & Schoenfield (25, 26) reported that the administration of CDCA (or chenic acid) decreased cholesterol saturation of bile in subjects with and without gallstones. These observations paved the way for the use of CDCA to dissolve gallstones. Danziger et al (27) reported the dissolution of human stones in January of 1972, followed in December of the same year by Bell et al (28). Subsequently there have been reports of larger series of patients treated in London (29), Bologna (30), Rochester (31), and Dundee (32), and a number of review papers are available (33–35).

Pharmacology of CDCA

CDCA is administered as the unconjugated bile acid and is available in a variety of preparations of high chemical purity but differing crystal morphology (36). It dissolves rapidly in the intestine, being completely absorbed in doses up to 400 mg whether taken with food or while fasting. The bile salt is effectively extracted by the hepatocyte; the first-pass clearance is about 62% (37). Hepatic extraction is constant in doses up to 400 mg and peripheral blood levels of unconjugated bile salts are low. Rachmilewitz & Saunders (38) incubated human jejunal and rectal biopsies with CDCA and suggested that some of the bile salt was metabolized to polar metabolites. However, van Berge-Henegouwen & Hofmann (37), studying peripheral blood radioactivity in patients two hours after the ingestion of [^{3}H]CDCA, found less than 5% with a mobility different from the unconjugated bile salt.

CDCA is conjugated in the liver to glycine or taurine and secreted into the bile. In the gut the conjugated CDCA may be reabsorbed unaltered. The conjugated CDCA is cleared by the liver even more effectively than the unconjugated bile salt, and only a slight elevation of glycine or taurine conjugates is detected in the peripheral blood. On the other hand, once in the gut the newly secreted conjugated CDCA may undergo deconjugation, and free bile salt joining the pool of ingested CDCA.

Some of the intestinal CDCA is not reabsorbed and is excreted in the stool and some undergoes transformation to the 7β-hydroxy epimer, ursodeoxycholic acid, but the majority of the unabsorbed CDCA undergoes 7α-dehydroxylation to lithocholic acid, which in turn undergoes an enterohepatic circulation similar to CDCA. Because lithocholic acid is the major metabolite formed when CDCA is ingested and because it is potentially hepatotoxic, its metabolism is of relevance and has been described in detail (39–41).

In normal man about one fifth of the lithocholate formed is absorbed in the terminal ileum and colon; the remainder is lost in the feces where it is a major bile salt. In man the important event in the metabolism of lithocholic acid is the sulfation of the conjugated salts that occurs at the 3 position. Sulfation takes place in the liver but can also occur in the kidney.

Sulfation alters the metabolism, excretion, and toxicity of lithocholic acid. The sulfated conjugates are water soluble and thus, in contrast to the other bile salts, are poorly absorbed by the gut and are lost in the feces. Sulfation decreases passive absorption in the jejunum and active absorption in the ileum. Only 25% of sulfated conjugates are conserved whereas about 40% of the unsulfated lithocholates undergo an enterohepatic circulation similar to CDCA and about 60% are conserved. The effect of sulfation is therefore to prevent the accumulation of lithocholic acid within the enterohepatic circulation.

DOSE CDCA is administered as a gelatin capsule containing 125 mg of the free bile salt, taken with meals. It was prescribed initially on an empirical basis in doses varying from 0.75 to 4.5 g (27). Subsequently the majority of authors have administered CDCA according to body weight, generally 12–15 mg/kg/day. Prescribing sufficient bile salt to reduce biliary cholesterol saturation while at the same time avoiding the main side effect, diarrhea, is difficult. Barbara et al (30) give the minimum effective dose as 7.8 mg/kg/day, but most other studies report 14–15 mg/kg/day for optimal efficiency (29, 31).

Our group, however, prescribes a fixed dose; we recommend 1000 mg daily (32, 42). This is the equivalent of 14–15 mg/kg/day for the majority of patients. The mean duration of therapy was greater (1.27 yr) on 750 mg daily than on 1000 mg (0.56 yr), and no patient on 500 mg daily has had complete gallstone dissolution.

In an attempt to define the optimum dose Mok et al (43) fed patients 250, 500, 750, and 1000 mg daily and concluded that a dose between 500 and 750 mg daily was effective and safe. But this was based on alterations in biliary lipid analysis and is at variance with our observations on the outcome of stone dissolution. Thistle et al (31) found that although some patients responded within a range of 4–12 mg/kg/day the majority of nonresponders received less than 12 mg/kg/day. Whereas only 38% of patients receiving less than 15 mg/kg/day responded, 83% had partial or complete gallstone dissolution when taking 15 mg or more/per kg/day. If judged by changes in biliary lipid composition and the proportion of CDCA in bile, a daily dose of 15 mg/kg should suffice for most nonobese patients. The ideal lowest dose has not been accurately defined (44, 45).

Duration of Therapy and Response Rate

A response (complete or partial dissolution) has been recorded only with radiolucent gallstones. It is exceptional for a radioopaque stone to disappear although an outer radiolucent component may dissolve.

The response rate of radiolucent stones varies from 47 to 64% (29, 31, 32), but these figures are misleading because they include both partial and complete dissolution over a period ranging from six months to four years and they include patients still in the process of dissolution. The percentage response might well be greater.

CDCA given in an optimal dose unsaturates bile in about three weeks (46). Stones usually begin to show radiological evidence of a reduction in size after 6 months and most stones are dissolved by 18 months. Lucent stones that have resisted therapy for 24 months seldom disappear.

Changes in Bile Composition

The essential effect of CDCA is to reduce the cholesterol saturation in bile. Exposure of the gallstone to bile unsaturated by cholesterol presumably enables cholesterol to leach out of the stone. Treatment with CDCA increases the body pool of the bile sale, and the increase is positively correlated with the amount absorbed (47). Similarly the amount of CDCA in bile increases (27, 36, 48) in direct proportion to dose (30, 48) until CDCA amounts to between 90 and 95% of biliary bile acids. As CDCA becomes the major bile salt, the proportion of cholate and deoxycholate progressively decreases, and the body pools of these two bile salts are reduced. The total bile salt pool represents the sum of these three bile salts, so overall the total bile salt pool is only slightly increased or remains unchanged by the feeding of CDCA (27, 36).

The ingestion of CDCA causes an increase in biliary lithocholic acid. An early study (27) found between 10 and 30% of the biliary bile salts to be monohydroxy but subsequent reports have given a much smaller proportion, seldom greater than 10% (32, 41). Increasing the dose from 750 to 1000 mg daily does not further increase biliary lithocholate (32). Most of the lithocholate is in the sulfated form (41) and although production and absorption of lithocholic acid increases, none of this bile salt is retained (49, 50).

When CDCA is administered, ursodeoxycholic acid (UDCA) appears in bile. Normally UDCA accounts for less than 1% of total biliary bile acids; this increases upon CDCA therapy to around 10% although it may be as much as 37% (48, 51). Why more UDCA forms remains unknown; it might relate to either bacterial transformation of CDCA in the gut or to direct epimerization of CDCA in the liver (51).

Most (25–27, 29, 30), but not all (32, 36, 48), patients receiving CDCA achieve unsaturation so that biliary lipid composition falls within the micellar zone. Biliary lipid analysis is of value when comparing groups, not individuals. In general, however, when CDCA amounts to some 70% or more of biliary bile acids, the bile will become unsaturated with cholesterol.

Northfield et al (52) measured the 24-hour biliary lipid output and composition during therapy with 1.0–2.5 g of CDCA daily. They demonstrated that unsaturation of bile resulted from decreased biliary cholesterol output. It could not be attributed to an increase of either bile salt or phospholipid. In another study (53) gallstone-free individuals received 500 mg of CDCA daily and the biliary cholesterol secretion, which was not abnormal, was reduced. This was attributed to a decrease in biliary cholesterol secretion in the absence of any increase in bile salt pool size or secretion. CDCA has no effect on body cholesterol pools (54).

CDCA decreases HMG-Co A reductase activity by 40%, which approximates levels found in individuals without gallstones. Cholesterol 7α-hydroxylase activity is also somewhat reduced. The current view is that CDCA acts by a direct effect on the rate-limiting enzyme for cholesterol biosynthesis (23, 55).

Factors Influencing Response

Stones less than 5 mm in diameter reportedly dissolve readily, whereas those greater than 15 mm dissolve appreciably slower (29, 31). Our experience, however, has been different and we have been unable to predict the speed of response from the size of the gallstone (32). Only patients with cholesterol stones should be treated; those with calcium-containing or pigment stones are resistant to therapy.

It is unclear why some patients with radiolucent stones fail to respond to an apparently adequate dose of CDCA. In some instances the stones are probably not primarily cholesterol in type. Large stones with smooth borders are predominantly cholesterol, but smaller radiolucent stones with irregular contours are often not—approximately 20% of radiolucent stones may be noncholesterol (56, 57).

There are also patients who do not seem to achieve cholesterol unsaturation in bile (32, 39, 57). In some this is probably because of extreme obesity, which appears to confer resistance to CDCA, although unsaturation may be achieved with doses up to 18.5 mg/kg/day (58). HMG-Co A reductase may vary either in activity or resistance to suppression, because in contrast to many patients with gallstones, cholesterol unsaturation in bile is easily produced if CDCA is fed to patients without gallstones (59).

Patients with nonfunctioning gallbladders are not recommended for treatment since the altered gallbladder function probably prevents circulation of unsaturated bile around the stone, and, of course, because it would be impossible to follow therapy by using cholecystography as a guide to stone dissolution. Ultrasonography may one day be used to monitor the presence and size of gallstones in these patients. Treatment of a nonfunctioning gallbladder with CDCA is not associated with a return of function. If oral cholecystography reveals nonfunction, surgery is indicated.

Prediction of Response

It has been claimed that the response to CDCA can be predicted from the radiolucency of the stone and the cholesterol content of bile (57), but this has not been the experience of many investigators (32, 48). At present the best index of response is the oral cholecystogram. Patients are required to undergo regular cholecystography usually at intervals of 6, 12, and 24 months (31, 32). Once the gallbladder appears free of stones the cholecystogram is repeated two to three months later and, if still normal, the stones are considered dissolved. It is reported that the analysis of fasting biliary lipids during therapy can predict the individual response to CDCA (29), but this has not been our experience (32, 39) or that of Hofmann et al (48). An improved cholesterol saturation in bile is associated with an increase in the proportion of biliary CDCA; but the magnitude of the CDCA content of bile is a poor means of distinguishing the responding from the nonresponding patient. The percentage of CDCA is probably only of help if assessing compliance (48). On balance we believe regular examination of gallbladder or hepatic bile by duodenal drainage is of little diagnostic advantage. There is no alternative to regular oral cholecystography.

ADJUVANTS TO CDCA It would be advantageous to enhance the effect of CDCA, thereby reducing the duration of therapy and affecting apparently nonresponsive stones.

A cholesterol-free diet prevented the anticipated decrease in biliary cholesterol saturation when CDCA was fed to hamsters (60), but contrary finding was reported in man where a low cholesterol diet (100 mg daily) enhanced the effect of CDCA (61). Results are conflicting when plant sterols are given at the same time as CDCA (31, 61), but it is generally agreed that there is no advantage to be gained by feeding them with the bile salt.

Phenobarbitone increases HMG-Co A reductase and 7α-hydroxylase activities (23) and increases the bile salt pool size, but it does not confer any therapeutic advantage when administered with CDCA (62).

Weight reduction might be advisable because obese patients are resistent to CDCA. Insufficient work has been undertaken on this problem but the observation that bile becomes more saturated with cholesterol during fasting (63) suggests that concomitant calorie restriction might not always be beneficial. Theoretically stone dissolution is enhanced in rapidly flowing bile (64) but this information has yet to be put to practical use.

Side Effects

The drug is well tolerated and many patients claim a sense of well-being with an improvement in abdominal discomfort and dyspepsia (29). Why this should be so is not clear and it is difficult to separate from a placebo effect.

The most important problem during therapy is diarrhea, which has been reported with varying frequency (31, 32, 37). It usually begins within the first few weeks of treatment, is dose related, and subsides after a few months. Most patients require only a reduction in the dose of CDCA and many will tolerate larger doses subsequently. Troublesome diarrhea prevents only very few patients from receiving CDCA and it is in these that UDCA is helpful. The diarrhea results from the stimulation of colonic mucosal adenyl cyclase and subsequent increase in net luminal secretion of fluid.

An occasional patient will complain of mild pruritus in the absence of biochemical cholestasis or markedly elevated serum bile salt concentrations.

A constant finding is elevated serum aspartate transaminase (SGOT) concentrations within one month of initiating therapy. The mean elevation is about 15%, usually within the normal range of serum concentration (29). It may be sustained for two years (32). The rise of enzyme is unrelated to an increase in serum lithocholate concentration (31). Gamma glutamyl transpeptidase concentrations fall to 44% of original values after two years of therapy. There is an immediate and as yet unexplained increase of blood glucose amounting to 15% of original values (32).

Animal studies suggested that CDCA might be hepatotoxic because of lithocholic acid formation (65, 66), but despite the serum transaminase elevation none of the studies of liver histology in man have shown any appreciable change (67–69). Some biopsies show "mild changes" consisting of a slight increase in fat and lipofunction. On the basis of electron microscopic findings Bateson et al (70) reported a significant increase in sinusoidal lipocytes, or Ito cells. The implications of this observation are uncertain, and liver function and histology should continue to be monitored in patients on long-term therapy.

The lack of hepatotoxicity of CDCA in man can be related to the sulfation of its main bacterial metabolite, lithocholic acid (39–41). Thus, in contrast to the rhesus monkey and the baboon, the increase of biliary lithocholate in man during therapy is between 1 and 3% (49, 50). It is relevant that the squirrel monkey, which also fails to accumulate lithocholate during CDCA administration, shows no evidence of hepatic damage (71). There is no major alteration in cholesterol or lipid metabolism (54) although a fall in serum triglycerides has been observed by some groups (29, 72, 73) but not by others (31, 74).

Another concern during the early days of therapy was that the partly dissolved gallstones would move into the common bile duct causing colic and jaundice. This is a most uncommon occurrence. Occasionally a patient develops acute cholecystitis while on therapy, but this is probably coincidence.

Indications and Contraindications

The indications for placing a patient on CDCA have yet to be established and most studies do not clearly explain why the patients were recommended for drug therapy in preference to cholecystectomy. Our practice is to offer the alternatives of medical or surgical treatment to patients with lucent stones and a functioning gallbladder. Alternatively CDCA may be reserved for those patients who reject surgery or who are regarded by the surgeons as unfit. The drug has been administered to women of childbearing age who are using adequate contraceptive techniques, and we have used it without difficulty in women on the oral contraceptive pill. The safety of CDCA administered during pregnancy is not established and it should be avoided in this circumstance. CDCA is inappropriate therapy for patients with severe gallstone symptoms because of the time taken to dissolve gallstones. It should not be used in icteric patients or in patients with nonfunctioning gallbladders.

Therapy is continued until the gallstones have dissolved. For patients whose gallstones do not dissolve, therapy is usually discontinued after two and a half to three years of nonresponse on the optimal dose. Other therapeutic end-points include the development of severe biliary symptoms, jaundice, a nonfunctioning gallbladder, or intolerable diarrhea.

Recurrence of Gallstones

Evidence is accumulating that gallstones may recur once the bile salt therapy has been stopped because the bile reverts to the supersaturated state. The natural history of gallstone disease after successful dissolution remains to be determined, but Thistle et al (31) reported that stones recurred in 3 of 15 patients followed for 48 months. Iser et al (29) recorded two recurrences in ten patients observed for 24 months. Readministration of CDCA induces rapid stone dissolution.

There are a number of management options once gallstones recur, and further studies are necessary to identify the best alternative for a particular situation. The physician can recommend cholecystectomy with its small but significant morbidity and mortality, or long-term CDCA therapy with its undetermined risks. Intermittent bile salt therapy might be considered, but it seems unlikely to be successful (46). Before long-term CDCA therapy can be widely recommended it will be necessary to define the lowest dose necessary to maintain unsaturated bile in patients whose stones have been dissolved. Some form of diet therapy is an attractive concept and further work is needed on the effects of low cholesterol or high roughage diets in this situation.

CDCA and Blood Lipids

In 1973 Bell et al (72) reported that elevated serum triglycerides fell to normal values while serum cholesterol remained unaltered in gallstone patients who had received CDCA. Miller & Nestel (73) recorded a 41% fall in serum triglycerides and a 10% drop in serum cholesterol. Our studies show that CDCA lowers elevated serum triglyceride levels by 10% and also initiates a modest fall in serum cholesterol concentrations (75). In contrast to the earlier studies (72) we (75) and others (30) have not observed a reduction of normal triglyceride levels.

Clofibrate is used extensively in the treatment of hyperlipidemia but it causes bile to become oversaturated with cholesterol (76). Such patients have an increased frequency of gallstones (77). These observations have led to the introduction of CDCA (750 mg daily) in combination with clofibrate when treating hyperlipidemia. This combination reverses the cholesterol saturation of bile induced by clofibrate, and the biliary lipids revert to a more normal pattern (78).

URSODEOXYCHOLIC ACID

UDCA is the 7β epimer of CDCA with hydroxyl groups in the 3α and 7β positions. The bile salt has been used as a chologogue in Japan, where its effectiveness in dissolving gallstones was first reported (79). UDCA affects the bile in a manner similar to CDCA (80). It is readily absorbed, enters the enterohepatic circulation to become the major bile salt, and inhibits the synthesis of CDCA. UDCA reduces HMG-Co A reductase activity by 30% when fed to rhesus monkeys, an equivalent effect to CDCA (81), and a reduction of similar magnitude has been reported in man (82). The overall effect is to decrease cholesterol saturation in bile. There is a rise in the molar proportion of total bile salts and the bile salt and lecithin: cholesterol ratio increases from 11.5 to 21.0 (80).

Deoxycholic acid and, more importantly, lithocholic acid virtually disappear from the bile (80). In a comparative study of CDCA and UDCA fed to rhesus monkeys no liver toxicity was observed with UDCA (81). Serum enzymes of liver origin remained within normal limits and there were none of the histological features typical of CDCA toxicity: periportal inflammation, bile duct proliferation, or hepatic necrosis. Why UDCA fails to cause an increase in lithocholate is unknown; it has been suggested that UDCA suppresses intestinal bacterial activity or that the bacteria which dehydroxylate the 7-hydroxy group are stereospecific (81).

The few studies available in man bear out the promise of this bile salt as a gallstone-dissolving agent. In one report gallstones decreased in size in 6 of 16 patients receiving 150 mg daily and in 2 of 15 patients on 600 mg daily. Bile became unsaturated in all patients receiving 150 mg daily (83). In another study of 74 patients treated with 450 mg daily the overall response was 43% (84). Small stones dissolved more readily than large (>10 mm), and, surprisingly, 2 of 22 patients with calcified stones had dissolution while some response was observed in a further 6 patients. If only patients with radiolucent stones were considered, the response was 52%, which was comparable to CDCA.

Maton et al (82) found a significant correlation between the dose of UDCA and the reduction in cholesterol saturation of bile. The average dose necessary to achieve unsaturation was 9.7 mg/kg/day, which suggested that UDCA might be effective at two-thirds the dose of CDCA; but the authors advise larger doses to be certain that the highest possible response rate will be obtained. Patients also experience improvement of dyspeptic symptoms while taking UDCA. Our studies with doses of 250 mg and 500 mg daily are equally encouraging (85).

Thus far there has been no complaint of diarrhea (82–85), no alteration in serum concentrations of liver enzymes, and no change in liver histology.

Whether UDCA will have a therapeutic effect at a lower dose than CDCA remains to be determined, and further dose-response studies and therapeutic trials are required. The bile acid is expensive and is not yet generally available.

CHOLIC ACID

The initial studies of Thistle & Schoenfield (25, 26) included administering cholic acid at doses of 500–1000 mg daily. In contrast to CDCA no significant changes were found in the bile salt and lecithin : cholesterol ratio. This observation has been confirmed many times since (31, 48, 53, 86), and all reports agree that oral cholic acid does not dissolve gallstones and has no place in their treatment.

Current opinion holds that where CDCA inhibits HMG-Co A reductase activity, cholic acid probably suppresses both enzymes equally (87). However, Carrella & Dietschy (88) reported greater suppression of cholesterol synthesis when rats were fed cholic acid than when CDCA was administered. Carrella & Dietschy argued that the influence of enhanced cholesterol absorption should also be considered. The influence of bile salt feeding on cholesterol metabolism is clearly complex and controversial but the observation that oral cholic acid cannot dissolve gallstones is unchallenged.

OTHER AGENTS

Phospholipids

It is tempting to suggest that feeding phospholipids might dissolve gallstones because of the importance of lecithin in maintaining biliary cholesterol in micellar solution (2). An early report suggested that feeding 10 g of sterile soybean "lecithin" caused a significant increase in the phospholipid content of hepatic bile (89). The improbable combination of cholic acid and lecithin has been claimed to be effective in two of six patients (90). The oral administration of β-glycerophosphate has been reported to increase the phospholipid content of human bile (91).

There is no enterohepatic circulation of intact lecithin molecules (92) although taking lecithin by mouth does contribute to the biliary lecithin pool (93). Biliary lecithin secretion is determined principally by the magnitude of lecithin synthesis, which is regulated by the amount of dietary choline (94). Thistle et al (31) fed patients 6 g of choline daily but were unable to achieve unsaturation of bile, nor did the choline enhance the therapeutic effect of CDCA. There is no therapy capable of influencing specifically the biliary phospholipid output.

Phenobarbital

In the rhesus monkey, phenobarbital administration increases the bile salt pool size and decreases the biliary cholesterol secretion (95), which suggests that it might be beneficial in patients with cholesterol gallstones. Studies in man have not been promising (62, 96). Feeding phenobarbital with CDCA is not of benefit (31, 97). At present there is no value in using phenobarbital either alone or in combination with bile salts.

The intravenous cholangiographic agent ioglycamide (Biligram®, Schering Chemicals Ltd.) has a choleretic effect and significantly lowers biliary output of cholesterol and phospholipids but leaves bile salt secretion unaltered (98). This information has not been put to practical use for gallstone dissolution. Bell & Doran (99) fed Rowachol®, a proprietary essential oil preparation, to 27 patients and reported that seven responded in six months including complete dissolution in three. Rowachol was said to lower the cholesterol saturation in bile by specifically inhibiting HMG-Co A reductase activity and reducing biliary cholesterol secretion.

Diet

It seems improbable that dietary manipulation alone will sufficiently reduce biliary cholesterol saturation to effect gallstone dissolution. A cholesterol-lowering diet, rich in polyunsaturated fatty acids, is associated with an increase in gallstone disease (100). Weight loss in obese subjects may be accompanied by increased biliary cholesterol saturation (63). The addition of fiber to the diet in the form of bran (the fiber-rich fraction of wheat) reduces the amount of secondary bile salt, deoxycholate, in bile (101). At the same time biliary CDCA increases. These observations led to the suggestion that the CDCA pool size is regulated by the amount of deoxycholic acid in the enterohepatic circulation (102) and it was suggested that bile unsaturation might be achieved by feeding bran. Recent studies have not confirmed that deoxycholic acid has either a specific effect on CDCA synthesis or improves cholesterol solubility (103–105).

Further work is required on the effect of dietary fiber on biliary lipids. At present there is insufficient evidence to warrant feeding a high fiber diet to achieve stone dissolution.

COMMON BILE DUCT STONES

Retained or newly-formed common bile duct stones can be treated by reoperation, by mechanical extraction via duodenoscopic sphincterostomy, or by drug therapy. The more traditional approach is reoperation, choledochostomy, and T-tube drainage. The dramatic proliferation of endoscopic techniques in gastroenterology has enabled common bile duct stones to be removed via an endoscopic papillotomy of the ampulla of Vater. Comprehensive reviews are available of the technique, its indications, and its complications (106, 107).

If the stones are discovered at the time of the postoperative cholangiographic examination, while the T-tube is still in place, they may be removed under radiological control by introducing probes or snares, or the common bile duct may be irrigated with a variety of solutions in an attempt to dissolve the stones. The technique ascribed to Burhenne or modications thereof has been much acclaimed (106, 108, 109). Ultrasonography has been used to fragment common bile duct stones (110).

Cholic Acid Infusion

In 1972 Way et al (111) described the successful dispersion of retained common bile duct stones by the infusion of sodium cholate via the T-tube. A solution of 100 mM sodium cholate in saline, buffered to pH 7.5 and sterilized by millipore filtration, is dripped into the duct at the rate of 30 ml/hr for up to ten days. The pressure in the bile duct is monitored and should not exceed 25–30 cm of solution. Response is assessed by a repeat T-tube cholangiogram after five days and again at the tenth day of infusion. These observations have been confirmed by others (112, 113).

Heparin Infusion

The strong negative electrokinetic potential of heparin suggested that it might be a useful agent to dissolve gallstones. In vitro studies, which unfortunately were not comparable, gave conflicting results (114–116). Successful dissolution has been reported following T-tube infusion of 25,000 USP units of heparin every eight hours (117). Sodium cholate (15–20 g) in one liter of saline alternating or in combination with 20,000 IU heparin in one liter of saline over 24 hours has been used successfully in 19 of 26 patients (118). We have had 5 successes in 15 patients receiving 100 mM sodium cholate, 2 successes in 5 patients receiving 100,000 IU heparin, and dissolution in 1 of 2 patients infused with physiological saline (119).

Other agents that have been infused include lecithin (116), quaternary amines (111), and the terpene *d*-limonene (120). We found that the infusion of monoglyceride solutions caused diarrhea.

Side effects with sodium cholate are common. Diarrhea can be most troublesome and it is usually necessary to administer cholestyramine orally during the bile salt infusion; it is not always effective. We have encountered biliary colic, cholangitis, biliary peritonitis, worsening of a biliary fistula, and sodium overload. Infusion therapy may fail if the stones are not primarily cholesterol, if the stones are too big or placed above the T-tube, or if the T-tube is too long. Nevertheless it is worth attempting local infusion therapy during the postoperative period before reexploring the common bile duct.

Oral Bile Salt Therapy

There is no point in infusing CDCA via a T-tube; it is likely to be no more effective than cholic acid and should be administered orally. There are a number of reports of common bile duct stones being treated by oral CDCA. Some success has been recorded (31, 42), but there is a high frequency of side effects such as jaundice, colic, and cholangitis that necessitate operation (28, 29, 32). We do not recommend this form of treatment unless there are unusual circumstances.

DECISION MAKING FOR RETAINED COMMON BILE DUCT STONES

There are a number of options depending upon the clinical situation and the availability of techniques. If stones are diagnosed when the T-tube is still present an attempt should be made to remove them mechanically via the fistula of the T-tube using a Burhenne catheter. Alternatively, local infusion therapy with either sodium cholate or heparin can be used. Failing this, endoscopic papillotomy may be tried; but the skilled endoscopist may rightly consider that an early approach to the papilla is preferable to prolonged infusion and hospitalization. Surgery always remains an effective option. If duct stones are discovered after removal of the T-tube, the choice lies between surgery and endoscopic papillotomy.

FUTURE PROSPECTS

Safe dissolution of cholesterol gallstones by oral therapy is now part of current medical practice. Unfortunately not all patients with radiolucent stones respond, and patients with calcium or pigment stones are unresponsive. We need to be able to identify with greater certainty prior to therapy which patients will respond, and a more practical method than repeated oral cholecystography is needed to study the progress of the gallstone dissolution. Almost certainly CDCA and UDCA will not be unique in dissolving stones and efforts should be directed toward obtaining more rapid and economical oral agents. Perhaps we should seek drugs capable of inhibiting crystal growth in saturated bile. It is essential to prevent gallstone recurrence and here dietary therapy has much attraction and potential. Agents are needed to dissolve calcium-containing and bile pigment gallstones.

Large programs for the study of gallstone disease will stimulate the acquisition of knowledge. In the US the National Cooperative Gallstone Study has been created primarily to determine the efficacy and safety of CDCA (121). Undoubtedly further extensive multicenter studies will be necessary before the ideal drug or drugs are evolved for the effective and safe dissolution of gallstones.

Literature Cited

1. Small, D. M., Bourgès, M., Dervichian, D. G. 1966. Ternary and quarternary aqueous systems containing bile salt, lecithin and cholesterol. *Nature (London)* 211:816–18
2. Niederhiser, D. H., Roth, H. P., Webster, L. T. Jr. 1966. Studies on the importance of lecithin for cholesterol solubilization in bile. *J. Lab. Clin. Med.* 68:90–102
3. Small, D. M. 1967. Physical-chemical studies of cholesterol gallstone formation. *Gastroenterology* 52:607–10
4. Carey, M. C., Small, D. M. 1978. The physical chemistry of cholesterol solubility in bile. *J. Clin. Invest.* 61:988–1026
5. Admirand, W. H., Small, D. M. 1968. The physical-chemical basis of cholesterol gallstone formation in man. *J. Clin. Invest.* 47:1043–52
6. Nakayama, F., van der Linden, W. 1970. Bile from gallbladder harbouring gallstone: Can it indicate stone formation? *Acta Chir. Scand.* 136:605–10
7. Smallwood, R. A., Jablonski, P., Watts, J. McK. 1972. Intermittent secretion of abnormal bile in patients with cholesterol gallstones. *Br. Med. J.* 4:263–66
8. Holzbach, R. T., Marsh, M., Olszewski, Holan K. 1973. Cholesterol solubility in bile. Evidence that supersaturated bile is frequent in healthy man. *J. Clin. Invest.* 52:1467–79
9. Small, D. M., Rapo, S. 1970. Source of abnormal bile in patients with cholesterol gallstones. *N. Engl. J. Med.* 283:53–57
10. Vlahcevic, Z. R., Bell, C. C. Jr., Swell, L. 1970. Significance of the liver in the production of lithogenic bile in man. *Gastroenterology* 59:62–69
11. Simmons, F., Ross, A. P. J., Bouchier, I. A. D. 1972. Alterations in hepatic bile composition after cholecystectomy. *Gastroenterology* 63:466–71
12. Vlahcevic, Z. R., Bell, C. C. Jr., Buhac, I., Farrar, J. T., Swell, L. 1970. Diminished bile acid pool size in patients with gallstones. *Gastroenterology* 59:165–73
13. Grundy, S. M., Metzger, A. L., Adler, R. D. 1972. Mechanism of lithogenic bile formation in American Indian women with cholesterol gallstones. *J. Clin. Invest.* 51:3026–43
14. Shaffer, E., Small, D. M. 1977. Biliary lipid secretion in cholesterol gallstone disease. *J. Clin. Invest.* 59:828–40
15. Northfield, T. C., Hofmann, A. F. 1973. Biliary lipid secretion in gallstone patients. *Lancet* 1:747–48
16. Mok, H. Y. I., von Bergmann, K., Grundy, S. M. 1977. Regulation of pool size of bile acids in man. *Gastroenterology* 73:684–90
17. Deleted in proof
18. Deleted in proof
19. Deleted in proof
20. Deleted in proof
21. Nicolau, G., Shefer, S., Salen, G. 1974. Determination of hepatic 3-hydroxy-3-methylglutaryl-CoA reductase activity in man. *J. Lipid Res.* 15:94–98
22. Salen, G., Nicolau, G., Shefer, S., Mosbach, E. H. 1975. Hepatic cholesterol metabolism in patients with gallstones. *Gastroenterology* 69:676–84
23. Coyne, M. J., Bonorris, G. G., Goldstein, L. I., Schoenfield, L. J. 1976. Effect of chenodeoxycholic acid and

phenobarbital on the rate-limiting enzymes of hepatic cholesterol and bile acid synthesis in patients with gallstones. *J. Lab. Clin. Med.* 87:281–91

24. Heller, F., Bouchier, I. A. D. 1973. Cholesterol and bile salt studies on the bile of patients with cholesterol gallstones. *Gut* 14:83–88
25. Thistle, J. L., Schoenfield, L. J. 1971. Lithogenic bile among young Indian women. Lithogenic potential decreased with chenodeoxycholic acid. *N. Engl. J. Med.* 284:177–81
26. Thistle, J. L., Schoenfield, L. J. 1971. Induced alterations in composition of bile of persons having cholelithiasis. *Gastroenterology* 61:488–96
27. Danziger, R. G., Hofmann, A. F., Schoenfield, J. L., Thistle, J. L. 1972. Dissolution of cholesterol gallstones by chenodeoxycholic acid. *N. Engl. J. Med.* 286:1–8
28. Bell, G. D., Whitney, B., Dowling, R. H. 1972. Gallstone dissolution in man using chenodeoxycholic acid. *Lancet* 2:1213–16
29. Iser, J. H., Dowling, R. H., Mok, H. Y. I., Bell, G. D. 1975. Chenodeoxycholic acid treatment of gallstones. A follow-up report and analysis of factors influencing response to therapy. *N. Engl. J. Med.* 293:378–83
30. Barbara, L., Roda, E., Roda, A., Sama, C., Festi, D., Mazzella, G., Aldini, R. 1976. The medical treatment of cholesterol gallstones: Experience with chenodeoxycholic acid. *Digestion* 14:209–19
31. Thistle, J. L., Hofmann, A. F., Orr, B. J., Stephens, D. H. 1978. Chenotherapy for gallstone dissolution. 1. Efficacy and safety. *J. Am. Med. Assoc.* 239:1041–46
32. Bateson, M. C., Ross, P. E., Murison, J., Bouchier, I. A. D. 1978. Comparison of fixed doses of chenodeoxycholic acid for gallstone dissolution. *Lancet* 1: 1112–14
33. Schoenfield, L. J. 1974. Medical therapy for gallstones. *Gastroenterology* 67: 725–29
34. Bell, G. D. 1974. The present position concerning gallstone dissolution. *Gut* 15:913–29
35. Bouchier, I. A. D. 1976. Gallstones. *Br. Med. J.* 2:870–72
36. van Berge-Henegouwen, G. P., Hofmann, A. F., Gaginella, T. S. 1977. Pharmacology of chenodeoxycholic acid. I. Pharmaceutical properties. *Gastroenterology* 73:291–99
37. van Berge-Henegouwen, G. P., Hofmann, A. F. 1977. Pharmacology of chenodeoxycholic acid. II. Absorption and metabolism. *Gastroenterology* 73: 300–9
38. Rachmilewitz, D., Saunders, D. R. 1976. Metabolism of chenodeoxycholate by intestinal mucosa. *Gastroenterology* 71:82–86
39. Cowen, A. E., Korman, M. G., Hofmann, A. F., Cass, O. W. 1975. Metabolism of lithocholate in healthy man. I. Biotransformation and biliary excretion of intravenously administered lithocholate, lithocholylglycine, and their sulfates. *Gastroenterology* 69:59–66
40. Cowen, A. E., Korman, M. G., Hofmann, A. F., Cass, O. W., Coffin, S. B. 1975. Metabolism of lithocholate in healthy man. II. Enterohepatic circulation. *Gastroenterology* 69:67–76
41. Cowen, A. E., Korman, M. G., Hofmann, A. F., Thomas, P. J. 1975. Metabolism of lithocholate in healthy man. III. Plasma disappearance of radioactivity after intravenous injection of labeled lithocholate and its derivatives. *Gastroenterology* 69:77–82
42. James, O., Cullen, J., Bouchier, I. A. D. 1975. Chenodeoxycholic acid therapy for gallstones: Effectiveness, toxicity and influence on bile acid metabolism. *Q. J. Med.* 44:349–67
43. Mok, H. Y. I., Bell, G. D., Dowling, R. H. 1974. Effect of different doses of chenodeoxycholic acid on bile-lipid composition and on frequency of side-effects in patients with gallstones. *Lancet* 2:253–57
44. Thistle, J. L., Hofmann, A. F., Yu, P. Y. S., Orr, B. 1977. Effect of varying doses of chenodeoxycholic acid on bile lipid and biliary bile acid composition in gallstone patients: A dose-response study. *Am. J. Dig. Dis.* 22:1–6
45. Marks, J. W., Bonorris, G. G., Chung, A., Coyne, M. J., Okun, R., Lachin, J. M., Schoenfield, L. J. 1977. Feasibility of low-dose and intermittent chenodeoxycholic acid therapy of gallstones. *Am. J. Dig. Dis.* 22:856–60
46. Iser, J. H., Murphy, G. M., Dowling, R. H. 1977. Speed of change in biliary lipids and bile acids with chenodeoxycholic acid—is intermittent therapy feasible? *Gut* 18:7–15
47. Danziger, R. G., Hofmann, A. F., Thistle, J. L., Schoenfield, L. J. 1973. Effect of oral chenodeoxycholic acid on bile acid kinetics and biliary lipid composition in women with cholelithiasis. *J. Clin. Invest.* 52:2809–21
48. Hofmann, A. F., Thistle, J. L., Klein, P. O., Szczepanik, P. A., Yu, P. Y. S. 1978. Chenotherapy for gallstone dissolution.

II. Induced changes in bile composition and gallstone response. *J. Am. Med. Assoc.* 239:1138–44
49. Allan, R. N., Thistle, J. L., Hofmann, A. F., Carter, J. A., Yu, P. Y. S. 1976. Lithocholate metabolism during chenotherapy for gallstone dissolution. 1. Serum levels of sulphated and unsulphated lithocholates. *Gut* 17:405–12
50. Allan, R. N., Thistle, J. L., Hofmann, A. F. 1976. Lithocholate metabolism during chenotherapy for gallstone dissolution. 2. Absorption and sulphation. *Gut* 17:413–19
51. Salen, G., Tint, G. S., Eliav, B., Deeling, N., Mosbach, E. H. 1974. Increased formation of ursodeoxycholic acid in patients treated with chenodeoxycholic acid. *J. Clin. Invest.* 53:612–21
52. Northfield, T. C., La Russo, N. F., Hofmann, A. F., Thistle, J. L. 1975. Biliary lipid output during three meals and an overnight fast. II. Effect of chenodeoxycholic acid treatment in gallstone subjects. *Gut* 16:12–17
53. Adler, R. D., Bennion, L. J., Duane, W. C., Grundy, S. M. 1975. Effects of low dose chenodeoxycholic acid feeding on biliary lipid metabolism. *Gastroenterology* 68:326–34
54. Hoffman, N. E., Hofmann, A. F., Thistle, J. L., 1974. Effect of bile acid feeding on cholesterol metabolism in gallstone patients. *Mayo Clin. Proc.* 49: 236–39
55. Goldstein, L. I., Bonorris, G. G., Coyne, M. J., Schoenfield, M. J. 1975. Persistent effects of chenodeoxycholic acid on biliary lipids in the hamster. *J. Lab. Clin. Med.* 85:1032–41
56. Bell, G. D., Dowling, R. H., Whitney, B. 1975. The value of radiology in predicting gallstone type when selecting patients for medical treatment. *Gut* 16:359–64
57. Bruunsgaard, A., Malver, E., Pedersen, L. R., Schlichting, P., Sylvest, J. 1977. Criteria for selection of patients for medical treatment (chenodeoxycholic acid therapy) of gallstones. *Scand. J. Gastroenterol.* 12:97–102
58. Iser, J. H., Maton, P. N., Murphy, G. M., Dowling, R. H. 1978. Resistance to chenodeoxycholic acid (CDCA) treatment in obese patients with gallstones. *Br. Med. J.* 1:1509–12
59. Bateson, M. C., Ross, P. E., Murison, J., Bouchier, I. A. D. 1977. Effect of prolonged feeding with chenodeoxycholic acid on bile in patients with and without gallstones. *Gut* 18:599–605
60. Coyne, M. J., Bonorris, G. G., Chung, A., Cove, H., Schoenfield, L. J. 1977. Dietary cholesterol affects chenodeoxycholic acid action on biliary lipids. *Gastroenterology* 72:927–31
61. Maudgal, D. P., Bird, R., Blackwood, W. S., Northfield, T. C. 1978. Low-cholesterol diet: enhancement of effect of CDCA in patients with gallstones. *Br. Med. J.* 2:851–53
62. Coyne, M. J., Bonorris, G. G., Chung, A., Goldstein, L. I., Lahara, D., Schoenfield, L. J. 1975. Treatment of gallstones with chenodeoxycholic acid and phenobarbital. *N. Engl. J. Med.* 292: 604–7
63. Bennion, L. J., Grundy, S. M. 1975. Effects of obesity and caloric intake on biliary lipid metabolism in man. *J. Clin. Invest.* 56:996–1011
64. Tao, J. C., Cussler, E. L., Evans, D. F. 1974. Accelerating gallstone dissolution. *Proc. Natl. Acad. Sci. USA* 71: 3917–21
65. Dyrszka, H., Salen, G., Zaki, F. G., Chen, T., Mosbach, E. H. 1976. Hepatic toxicity in the Rhesus monkey treated with chenodeoxycholic acid for 6 months: biochemical and ultrastructural studies. *Gastroenterology* 70:93–104
66. McSherry, C. K., Morrissey, K. P., Swarm, R. L., May, P. S., Niemann, W. H., Glenn, F. 1976. Chenodeoxycholic acid induced liver injury in pregnant and neonatal baboons. *Ann. Surg.* 184:490–99
67. James, O. F. W., Scheuer, P. J., Bouchier, I. A. D. 1973. The effect of chenodeoxycholic acid on liver function and histology in man. *Digestion* 8:432
68. Bell, G. D., Mok, H. Y. I., Thare, M., Murphy, G. M., Henry, K., Dowling, R. H. 1974. Liver structure and function in cholelithiasis: effect of chenodeoxycholic acid. *Gut* 15:165–72
69. Pederson, L., Bremmelgaard, A. 1976. Hepatic morphology and bile acid composition of bile and urine during chenodeoxycholic acid therapy for radiolucent gallstones. *Scand. J. Gastroenterol.* 11:385–89
70. Bateson, M. C., Hopwood, D., Bouchier, I. A. D. 1977. Effect of gallstone-dissolution therapy on human liver structure. *Am. J. Dig. Dis.* 22:293–99
71. Suzaki, H., Yoshida, T., Miki, H., Sado, T., Hashimoto, K. 1977. Lack of toxicity of chenodeoxycholic acid in the squirrel monkey. *Gastroenterology* 73: 310–13

72. Bell, G. D., Lewis, B., Petrie, A., Dowling, R. H. 1973. Serum lipids in cholelithiasis: Effect of chenodeoxycholic acid therapy. *Br. Med. J.* 3:520–22
73. Miller, N. E., Nestel, P. J. 1974. Triglyceride lowering effect of chenodeoxycholic acid in patients with endogenous hypertriglyceridaemia. *Lancet* 2:929–31
74. Schlierf, G., Stiehl, A., Heuck, C. C., Larg, P. D., Oster, P., Schellenberg, B. 1976. Does chenodeoxycholic acid have a lipid lowering effect in primary hyperlipoproteinemia? *Eur. J. Clin. Pharmacol.* 10:147–49
75. Bateson, M. C., Maclean, D., Evans, J. R., Bouchier, I. A. D. 1978. Chenodeoxycholic acid therapy for hypertriglyceridaemia in men. *Br. J. Clin. Pharmacol.* 5:249–54
76. Pertsemlidis, D., Panveliwalla, D., Ahrens, E. H. Jr. 1974. Effects of clofibrate and of an estrogen-progestin combination on fasting biliary lipids and cholic acid kinetics in man. *Gastroenterology* 66:565–73
77. Bateson, M. C., Maclean, D., Ross, P. E., Bouchier, I. A. D. 1978. Clofibrate therapy and gallstone induction. *Am. J. Dig. Dis.* 23:623–28
78. Bateson, M. C., Ross, P. E., Murison, J., Bouchier, I. A. D. 1978. Reversal of clofibrate-induced cholesterol oversaturation of bile with chenodeoxycholic acid. *Br. Med. J.* 1:1171–73
79. Makino, I., Shinozaki, K., Yoshino, K., Nakagawa, S. 1975. Dissolution of cholesterol gallstones by ursodeoxycholic acid. *Jpn. J. Gastroenterol.* 72:690–702
80. Fedorowski, T., Salen, G., Colallilo, A., Tint, G. S., Mosbach, E. H., Hall, J. C. 1977. Metabolism of ursodeoxycholic acid in man. *Gastroenterology* 73: 1131–37
81. Fedorowski, T., Salen, G., Zaki, F. G., Shefer, S., Mosbach, E. H. 1978. Comparative effects of ursodeoxycholic acid and chenodeoxycholic acid in the rhesus monkey. *Gastroenterology* 74: 75–81
82. Maton, P. N., Murphy, G. M., Dowling, R. H. 1977. Ursodeoxycholic acid treatment of gallstones. *Lancet* 2:1297–1301
83. Nakagawa, S., Makino, I., Ishizaki, T., Dohi, I. 1977. Dissolution of cholesterol gallstones by ursodeoxycholic acid. *Lancet* 2:367–69
84. Okumura, M., Tanikawa, K., Chuman, Y., Koji, T., Nakagawa, S., Nakamura, Y., Iino, H., Yamasaki, S., Hisatsugu, T. 1977. Clinical studies on dissolution of gallstones using ursodeoxycholic acid. *Gastroenterol. Jpn.* 12:469–75
85. Bateson, M. C., Ross, P. E., Murison, J., Bouchier, I. A. D. 1978. Effects of ursodeoxycholic acid treatment (UDCA) on biliary lipids in man. *Gut* 19:A972
86. La Russo, N. F., Hoffman, N. E., Hofmann, A. F., Northfield, T. C., Thistle, J. L. 1975. Effect of primary bile acid ingestion on bile acid metabolism and biliary lipid secretion in gallstone patients. *Gastroenterology* 69:1301–14
87. Schoenfield, L. J., Bonorris, G. G., Ganz, P. 1973. Induced alterations in the rate-limiting enzymes of hepatic cholesterol and bile acid synthesis in the hamster. *J. Lab. Clin. Med.* 82:858–68
88. Carrella, M., Dietschy, J. M. 1977. Comparison of the effects of cholic acid and chenic acid feeding on rates of cholesterol synthesis in the liver of the rat. *Am. J. Dig. Dis.* 22:318–26
89. Tompkins, R. K., Barke, L. G., Zollinger, R. M., Cornwell, D. G. 1970. Relationship of biliary phospholipid and cholesterol concentrations to the occurrence and dissolution of human gallstones. *Ann. Surg.* 172:936–45
90. Touli, J., Jablonski, P., Watts, J. McK. 1975. Gallstone dissolution in man using cholic acid and lecithin. *Lancet* 2:1124–26
91. Linscheer, W. G., Raheja, K. L. 1974. Effect of glycerophosphate on lithogenic bile. *Lancet* 2:551–53
92. Saunders, D. R. 1970. Insignificance of the enterobiliary circulation of lecithin in man. *Gastroenterology* 59:848–52
93. Tompkins, R. K., King, W. III 1974. Investigations of the enterobiliary metabolism of lecithin. *Surgery* 75:243–52
94. Robins, S. J., Armstrong, M. J. 1976. Biliary lecithin secretion. II. Effects of dietary choline and biliary lecithin synthesis. *Gastroenterology* 70:397–402
95. Redinger, R. N., Small, D. M. 1973. Primate biliary physiology. VIII. The effect of phenobarbital upon bile salt synthesis and pool size, biliary lipid secretion, and bile composition. *J. Clin. Invest.* 52:161–72
96. Hepner, G. W. 1976. The effect of phenobarbital on biliary lipid metabolism and hepatic microsomal drug metabolism in patients with cholesterol cholelithiasis. *Am. J. Dig. Dis.* 21:370–75
97. Ginsberg, R. L., Garnick, M. B. 1977. Effect of phenobarbital on biliary lipid metabolism in normal man. *Gastroenterology* 72:1221–27

98. Bell, G. D., Doran, J., Fayadh, M., Murphy, G., Dowling, R. H. 1978. Effect of ioglycamide (Biligram) on bile flow and biliary lipid secretion in man. *Gut* 19:300–7
99. Bell, G. D., Doran, J. 1979. Gallstone dissolution in man using an essential oil preparation. *Br. Med. J.* 1:29
100. Sturdevant, R. A. L., Perru, M. L., Dayton, S. 1973. Increased prevalence of cholelithiasis in men ingesting a serum-cholesterol-lowering diet. *N. Engl. J. Med.* 288:24–27
101. Pomare, E. W., Heaton, K. W. 1973. Alteration of bile salt metabolism by dietary fibre (bran). *Br. Med. J.* 4: 262–64
102. Pomare, E. W., Low-Beer, T. S. 1975. The selective inhibition of chenodeoxycholate synthesis by cholate metabolites in man. *Clin. Sci. Mol. Med.* 48:315–21
103. Wicks, A. C. B., Yeats, J., Heaton, K. W. 1978. Bran and bile: time-course of changes in normal young men given a standard dose. *Scand. J. Gastroenterol.* 13:289–92
104. Ahlberg, J., Angelin, B., Einarsson, K., Hellström, K., Leijd, B. 1977. Influence of deoxycholic acid on biliary lipids in man. *Clin. Sci. Mol. Med.* 53:249–56
105. La Russo, N. F., Szczepanik, P. A., Hofmann, A. F. 1977. Effect of deoxycholic acid ingestion on bile acid metabolism and biliary lipid secretion in normal subjects. *Gastroenterology* 72: 132–40
106. Classen, M., Ossenberg, F. W. 1977. Non-surgical removal of common bile duct stones. *Gut* 18:760–69
107. Safrany, L. 1977. Duodenoscopic sphincterotomy and gallstone removal. *Gastroenterology* 72:338–43
108. Burhenne, H. J. 1973. Non-operative retained biliary tract stone extraction. A new roentgenologic technique. *Am. J. Roentgenol.* 117:388–99
109. Garrow, D. G. 1977. The removal of retained biliary tract stones: report of 105 cases. *Br. J. Radiol.* 50:777–82
110. Davies, H., Bean, W. J., Barnes, F. S. 1977. Breaking up residual gallstones with an ultrasonic drill. *Lancet* 2:278–9
111. Way, L. W., Admirand, W. H., Dunphy, J. E. 1972. Management of choledocholithiasis. *Ann. Surg.* 176: 347–59
112. Lansford, C., Mehta, S., Kern, F. Jr. 1974. The treatment of retained stones in the common bile duct with sodium cholate infusion. *Gut* 15:48–51
113. Britton, D. C., Gill, B. S., Taylor, R. M. R., James, O. 1975. The removal of retained gallstones from the common bile duct: experience with sodium cholate infusion and the Burhenne catheter. *Br. J. Surg.* 62:520–23
114. Lahana, D. A., Bonorris, D. A., Schoenfield, L. J. 1974. Gallstone dissolution in vitro by bile acids, heparin and quaternary amines. *Surg. Gynecol. Obstet.* 138:683–85
115. Romero, R., Butterfield, W. C. 1974. Heparin and gallstones. *Am. J. Surg.* 127:687–88
116. Cheung, L. Y., Englent, E. Jr., Moody, F. G., Wales, E. E. 1974. Dissolution of gallstones with bile salts, lecithin, and heparin. *Surgery* 76:500–3
117. Gardner, B., Dennis, C. R., Patti, J. 1975. Current status of heparin dissolution of gallstones. *Am. J. Surg.* 130:293–95
118. Christiansen, L. A., Scherstén, T., Burcharth, F., Bruusgaard, A., Lindblad, L., Cahlin, E. 1977. Treatment of retained bile duct calculi with T-tube infusion of sodium cholate and heparin. *Scand. J. Gastroenterol.* 12:337–39
119. Bateson, M. C., Bouchier, I. A. D. 1977. Local infusion therapy for radiolucent retained common duct stones using sodium cholate, heparinised saline and physiological saline. *Gut* 18:A977
120. Igimi, H., Hisatsugu, T., Nishimara, M. 1976. The use of d-limonene preparation as a dissolving agent of gallstones. *Am. J. Dig. Dis.* 21:926–39
121. Schoenfield, L. J. 1977. Stone dissolution and the National Cooperative Gallstone Study. *Am. J. Dig. Dis.* 22: 1115–16

Ann. Rev. Med. 1980. 31:79–87
Copyright © 1980 by Annual Reviews Inc. All rights reserved

INTESTINAL ABSORPTION AND MALABSORPTION OF FOLATES

◆7344

Charles H. Halsted, M.D[1]

Department of Internal Medicine, University of California,
Davis, California 95616

INTRODUCTION

Folates are essential to a variety of reactions involved in nucleoprotein synthesis and cell turnover. The most common sign of folate deficiency is megaloblastic anemia; intestinal mucosal cell function may be less obviously affected. The minimal adult requirement for folate is 50 μg per day and the current recommended dietary allowance is 300 μg per day (1). Dietary folates exist predominantly as reduced and conjugated pteroylpolyglutamates with various substitutions on the pteridine ring. As recently reviewed (2), the process of folate absorption requires hydrolysis to pteroylmonoglutamate forms by an intestinal mucosal γ-carboxypeptidase (folate conjugase), and subsequent transport across the intestinal epithelium by a mechanism which probably involves one or more metabolic conversions. Most data on folate absorption has been obtained by using the more stable and available oxidized forms of pteroylpolyglutamate (PteGlu$_n$) and pteroylmonoglutamate (PteGlu).

Hydrolysis and Absorption of PteGlu$_n$

A series of studies at the University of Alabama showed that the absorption of synthetic ^{14}C–PteGlu$_n$ required its initial intestinal cleavage to ^{14}C–PteGlu. In studies of patient volunteers, the oral administration of ^{14}C–PteGlu$_n$ was followed by parallel serum rises in ^{14}C and folate activity only when the label was positioned in synthesis on the first glutamyl unit (3). A subsequent in vivo study placed ^{14}C–PteGlu$_n$ with more than three glutamyl units in an isolated dog intestinal segment and demonstrated that ^{14}C–PteGlu was the only labeled folate compound recoverable in the mesenteric venous blood (4). More recent data from the same model suggested

---

[1] Supported by USPHS grant AM 18330-05.

0066-4219/80/0401-0079$01.00

that the hydrolytic process of $PteGlu_n$ was rate limiting to its absorption, since the appearance of ^3H in the mesenteric venous plasma was more than twice that of ^{14}C after equal amounts and counts per minute of ^3H–PteGlu and ^{14}C–$PteGlu_5$ were placed in the intestinal loop (5).

To study folate absorption in humans, we used jejunal perfusion to compare the rates of absorption of ^{14}C–$PteGlu_7$ (the label positioned on the first glutamyl unit), with that of ^3H–PteGlu, both compounds delivered in equal amounts in the same saline solution. Absorption was expressed as percent luminal disappearance of each compound per 30 cm of jejunum (6) and as percent recovery of each label in a 48-hour urine collection, using a post-perfusion flushing dose of nonradioactive PteGlu (7). By either method, the absorption of ^3H–PteGlu was significantly greater than that of ^{14}C–$PteGlu_7$ (Table 1). Godwin & Rosenberg (8) obtained similar data by measuring the urinary recovery of ^3H after sequential oral administration of ^3H–PteGlu and ^3H–$PteGlu_7$. Our perfusion studies also permitted measurement of the in vivo hydrolysis of ^{14}C–$PteGlu_7$ by chromatographic detection of shorter chain length ^{14}C–$PteGlu_n$ residues within intestinal aspirates obtained downstream from the infusion site (9). Chromatograms such as that shown in Figure 1 suggested that the exposure of ^{14}C–$PteGlu_7$ to the mucosal surface during its perfusion resulted in its stepwise hydrolysis to ^{14}C–PteGlu. This concept was supported by in vitro experiments in which similar hydrolytic products were identified after incubation of ^{14}C–$PteGlu_7$ with homogenates of human intestinal mucosa but not after incubation with fasting intestinal luminal contents (6, 9).

Table 1 Jejunal perfusion of labeled folates in normal subjects and in patients with celiac or tropical sprue[a].

		Luminal disappearance (% per 30 cm)			Urinary recovery (% per 48 hr)	
	n	^3H–PteGlu	^{14}C–$PteGlu_7$	*n*	^3H	^{14}C
Normal	9	74.7±3.2	47.7±3.6	6	58.6±2.5	34.8±2.8
Celiac sprue						
untreated	6	8.8±2.7	4.5±2.2	4	35.5±4.7	16.3±4.9
treated	6	31.6±2.7	19.1±1.1	4	60.0±4.7	35.5±4.2
Tropical sprue						
untreated	6	25.4±2.9	20.7±2.2	6	35.7±4.5	16.3±3.0
treated	6	59.7±7.3	47.9±5.6	6	59.4±4.2	36.0±1.7

[a] Jenunal perfusion solutions were isotonic saline containing equimolar concentrations of ^3H–PteGlu and ^{14}C–$PteGlu_7$ (2 μM in normal and celiac sprue studies, 0.75 μM in tropical sprue studies). In each group the absorption of ^3H–PteGlu significantly exceeded that of ^{14}C–$PteGlu_7$, when expressed either as luminal disappearance or as urinary isotope recovery. Appropriate treatment of celiac or tropical sprue resulted in significant improvement of absorption of each labeled folate. Data summarized from References 6, 7, 9, and 40.

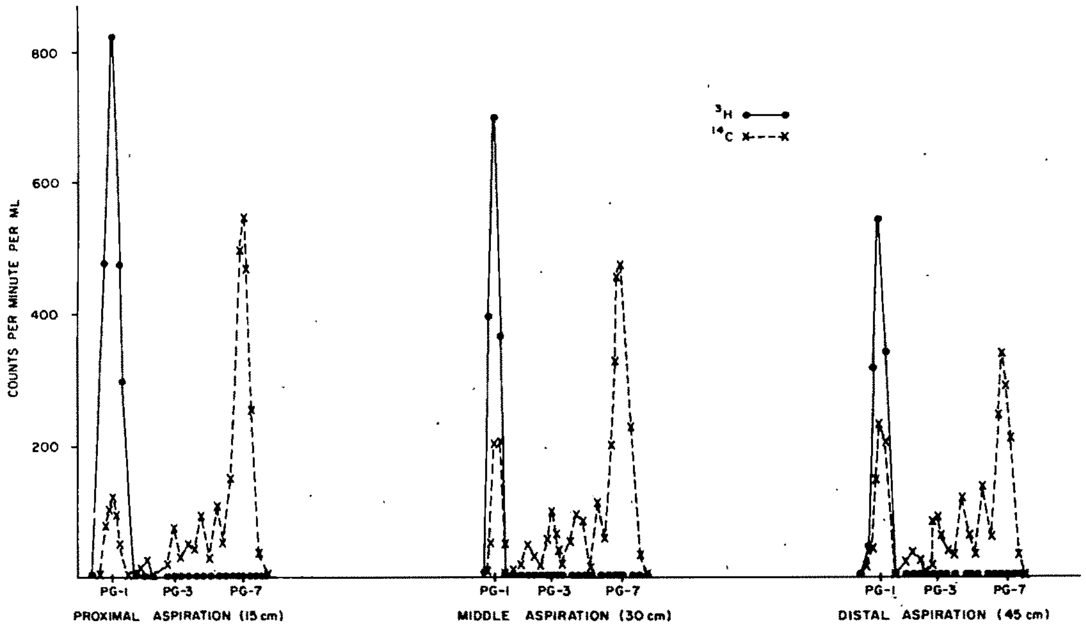


Figure 1. Chromatograms of intestinal aspirates obtained at three aspiration ports during jejunal perfusion at the ligament of Treitz of a normal subject using a saline solution with equimolar concentrations of ^{3}H–PteGlu and ^{14}C–PteGlu$_7$. The patterns show a progressive decrease in ^{14}C–PteGlu$_7$ and simultaneous increase in end-product ^{14}C–PteGlu, suggesting a process of stepwise hydrolysis on the surface of the intestinal mucosa. Reprinted from Reference 9 with permission of the American Gastroenterological Association.

Folate Conjugase

In comparison to its activity in bile, gastric, or pancreatic juice, the hydrolytic enzyme folate conjugase is concentrated several hundred-fold in the jejunal and ileal mucosa (10). In the chicken, purified intestinal folate conjugase is an endopeptidase (11). Several groups described guinea pig or rat intestinal folate conjugase as a lysosomal enzyme with a pH optimum of 4.5 (12–14). Recent studies from our laboratory defined two separate folate conjugase activities in human jejunal mucosa: one located on the brush border surface with a neutral pH optimum and the other located within the cell with an acid pH optimum (15). In the context of our jejunal perfusion studies (Figure 1), the brush border enzyme may play a critical role in the hydrolysis of PteGlu_n.

Absorption of PteGlu

Human intestinal perfusion studies showed that the absorption of PteGlu is maximal in the jejunum but negligible in the ileum (16), is enhanced by the presence of glucose (17), and is decreased by alklalinization of the perfusion solution (18). Since PteGlu is weakly acidic, its absorption might be enhanced by lowering pH, thus increasing nonpolarity and membrane diffusability (19). However, because of the presence of several dissociation constants (20, 21), PteGlu is nonpolar over a very narrow pH range between 6.0 and 6.5, and its uptake from the perfused rat jejunum, maximal at pH 6.0, is decreased both by acidification and alkalinization of the perfusate (20).

Other researchers described a binding protein on the brush border surface (22). The overall transport of PteGlu is best described as a two-phase system, with saturable (zero-order) kinetics below and linear (first-order) kinetics above a PteGlu concentration of about 10 μM (23, 24). "Saturability" may in fact be largely related to metabolic conversion of the molecule during its transport. Several groups have shown that reduction, methylation (25, 26), and formylation (27) occur during in vitro transport at low concentrations, whereas PteGlu is recovered unchanged from the human umbilical (portal) vein after oral administration of larger pharmacologic doses of the vitamin (28). The description of several infants with megaloblastic anemia responding to parenteral but not oral PteGlu (29, 30) is strong evidence for genetic control over a specific transport process.

FOLATE MALABSORPTION

Drugs or mucosal disease can induce folate malabsorption by decreasing the hydrolysis of PteGlu_n or the transport of PteGlu. Several authors place considerable emphasis on the potential effect of altered intraluminal pH on PteGlu absorption. Clinical conditions in which malabsorption contributes to folate deficiency include celiac and tropical sprue, and the chronic use or abuse of certain drugs including alcohol, diphenylhydantoin (Dilantin®), and Azulfidine®. Although the chronic use of oral contraceptives may induce changes in morphology suggestive of altered cellular folate metabolism, the serum folate level is normal (31, 32), as is the absorption of $PteGlu_n$ and the activity of jejunal folate conjugase (32).

Sprue

Celiac and tropical sprue are each characterized by generalized malabsorption and the common clinical presentation of megaloblastic anemia due to folate deficiency (33, 34). The gut lesion of celiac sprue occurs mainly in the upper intestine and is characterized by flattening of the villi and hyperplasia of the crypt epithelium, which responds in most cases to exclusion of gluten from the diet. On the other hand, the blunted or flattened villi of tropical sprue are associated with decreased mitoses and nuclear enlargement in the regenerative crypt epithelium (35). In most patients with tropical sprue the malabsorption syndrome and gut histology are corrected by combined treatment with folic acid and antibiotics.

Folate malabsorption was documented in patients with celiac or tropical sprue in studies that measured serum levels after an oral dose (36, 37) or the disappearance of labeled folate from the perfused jejunum (6, 38–40). A proposal that the malabsorption of PteGlu in celiac sprue is due to the injured mucosa's inability to maintain an optimal acid microclimate for absorption (19) was supported by the recent finding of small but significant increases in surface pH in mucosal biopsy specimens obtained from untreated patients (41). However, MacKenzie & Russell (18) found no correlation between decreased jejunal uptake of PteGlu and unchanged intraluminal pH in their perfusion studies of celiac sprue patients. We found significantly decreased jejunal uptake of perfused ^3H–PteGlu and ^{14}C–$PteGlu_7$ in untreated celiac and tropical sprue patients as compared to results obtained in the same patients after treatment (6, 40) (Table 1) even though intraluminal pH was neutral or slightly acidic. In the celiac sprue patients, distal small bowel adaptation to folate absorption was suggested by finding a disproportionate excretion of each label in the urine (a measure of total intestinal absorption) as compared to the calculated jejunal uptake from the perfused jejunal segment (7).

Alcoholism

Folate deficiency, defined as megaloblastic anemia with a low serum folate level, occurred in 40 and 61% of derelict alcoholic patients studied at three different municipal hospitals, but was less frequent in well-to-do alcoholics (42). Potential causes of folate deficiency in alcoholism include poor diet, intestinal malabsorption, and defective metabolism due to associated liver disease. Malabsorption of ^{3}H–PteGlu was documented in two studies of recently drinking malnourished alcoholics (43, 44). In a prospective study, decreased jejunal uptake of ^{3}H–PteGlu was induced in patient volunteers by feeding a folate-deficient diet to which was added alcohol equivalent to a daily pint of whiskey (45). This study suggested that PteGlu malabsorption in chronic alcoholism results from a synergistic effect of folate deprivation and ethanol toxicity on transport processes in the mucosa. To date, there are no studies of the effect of chronic or acute alcoholism on the absorption of PteGlu$_n$ or on the activity of intestinal folate conjugase. By contrast, a recent study of nondrinking alcoholic patients with pancreatic insufficiency suggested that the absorption of ^{3}H–PteGlu, indirectly measured by urine isotope recovery after an oral dose, was enhanced because of alkaline intraluminal pH (46).

Diphenylhydantoin (Dilantin) Therapy

Folate deficiency occurs in up to two thirds of epileptic patients who chronically receive Dilantin (47). An inhibitory effect of chronic Dilantin use on the absorption of PteGlu was suggested by several studies in drug-treated epileptics showing a less than normal rise in serum folate after an oral dose of PteGlu (48, 49). Benn et al (49) ascribed PteGlu malabsorption to intraluminal alkalinization, since duodenal pH was elevated in their patients, and since in normal subjects the serum folate rise could be acutely inhibited by either 300 mg of Dilantin or 10 g of oral bicarbonate (49). However, using the same method, Perry & Chanarin (50) found no acute effect of a similar dose of Dilantin or bicarbonate on the absorption of PteGlu or PteGlu$_n$ folate. Other evidence fails to support the intraluminal alkalinizing effect of Dilantin (51). Gerson et al (52) showed that the jejunal uptake of ^{3}H–PteGlu in normal volunteers could be suppressed by the addition of Dilantin to the perfusion solution in a concentration known to occur intraluminally after a single tablet, but not sufficient to change the pH of the perfusate. The effect of Dilantin on the hydrolysis and absorption of PteGlu$_n$ is less well understood. Two studies purporting to show Dilantin inhibition of intestinal folate conjugase (53, 54) were not confirmed by another (55).

Salicylazosulfapyridine (Azulfidine)

Folate deficiency in inflammatory bowel disease (Crohn's disease and ulcerative colitis) results partly from intestinal malabsorption, which may be specifically aggravated by the therapeutic agent Azulfidine. Franklin & Rosenberg (56) found significantly less urinary isotope excretion after oral ^{3}H–PteGlu in patients receiving Azulfidine that in those who were not, and that a single 1-g dose of Azulfidine decreased the serum folate rise after administration of oral PteGlu to normal subjects. In vitro, azulfidine, but not its metabolites sulfapyridine or 5-amino-salicylic acid, inhibited the uptake of ^{3}H–PteGlu by rat everted intestinal rings (56). The mechanism of Azulfidine suppression of PteGlu absorption might relate to altered metabolism, since the same group recently showed that activities of several hepatic folate-dependent enzymes could be competitively inhibited by Azulfidine in concentrations known to impair the intestinal transport of ^{3}H–PteGlu (57).

SUMMARY

Folates, which appear in the diet principally as PteGlu$_n$ forms, are hydrolyzed to the PteGlu form by intestinal mucosal folate conjugase. Evidence based on jejunal perfusion studies and a recent demonstration of brush border folate conjugase activity suggests that the hydrolytic step occurs on the mucosal surface prior to entry of PteGlu into the absorbing cell. The normal absorption of PteGlu, maximal in the jejunum, is affected by intraluminal pH and may require the sequential steps of membrane binding and transport, followed by intracellular reduction and methylation, prior to its appearance in the portal blood. Although convincing data indicate that folate absorption is decreased in diseases of the intestinal mucosa, and after the chronic administration of certain drugs, mechanisms of folate malabsorption are incompletely understood. For example, proof of the intriguing hypothesis that folate absorption is affected by diseases or drugs that alter intraluminal pH requires experimental design which measures pH and absorption simultaneously. Results from the jejunal perfusion method apparently disproved this association in celiac sprue patients (6, 18) and in normal subjects administered Dilantin (52). Greater understanding of the mechanisms of folate malabsorption will follow more precise definition of the effects of diseases and drugs on pH and on the various enzyme systems required for folate digestion and transport.

Literature Cited

1. Herbert, V. 1977. Summary of the workshop. In *Folic Acid*, ed. H. P. Broquist, C. E. Butterworth, C. Wagner, pp. 277–93. Washington, DC: Nat. Acad. Sci. 298 pp.
2. Halsted, C. H. 1979. The intestinal absorption of folates. *Am. J. Clin. Nutr.* 32:846–55
3. Butterworth, C. E. Jr., Baugh, C. M., Krumdieck, C. L. 1969. A study of folate absorption and metabolism in man utilizing carbon 14–labeled polyglutamates synthesized by the solid phase method. *J. Clin. Invest.* 48:1131–42
4. Baugh, C. M., Krumdieck, C. L., Baker, H. J., Butterworth, C. E. Jr. 1971. Studies on the absorption and metabolism of folic acid. Folate absorption in the dog after exposure of isolated intestinal segments to synthetic pteroylpolyglutamates of various chain lengths. *J. Clin. Invest.* 50:2009–21
5. Baugh, C. M., Krumdieck, C. L., Baker, H. J., Butterworth, C. E. Jr. 1975. Absorption of folic acid poly-

γ-glutamates in dogs. *J. Nutr.* 105: 80–89

6. Halsted, C. H., Reisenauer, A. M., Romero, J. J., Ruebner, B. 1977. Jejunal perfusion of simple and conjugated folates in celiac sprue. *J. Clin. Invest.* 59:933–40
7. Halsted, C. H., Reisenauer, A. M., Shane, B., Tamura, T. 1978. Availability of monoglutamyl and polyglutamyl folates in normal subjects and in patients with celiac sprue. *Gut* 19:886–91
8. Godwin, H. A., Rosenberg, I. H. 1975. Comparative studies of the intestinal absorption of ^{3}H–pteroylmonoglutamate and ^{3}H–pteroylheptaglutamate in man. *Gastroenterology* 69:364–73
9. Halsted, C. H., Baugh, C. M., Butterworth, C. E. Jr. 1975. Jejunal perfusion of simple and conjugated folates in man. *Gastroenterology* 68:261–69
10. Hoffbrand, A. V., Peters, T. J. 1970. Recent advances in knowledge of clinical and biochemical aspects of folate. *Schweiz. Med. Wochenschr.* 100: 1954–60
11. Rosenberg, I. H., Neumann, H. 1974. Multistep mechanism in the hydrolysis of pteroylpolyglutamate by chicken intestine. *J. Biol. Chem.* 249:5126–30
12. Hoffbrand, A. V., Peters, T. J. 1969. The subcellular localization of pteroylglutamate hydrolase and folate in guinea pig intestinal mucosa. *Biochim. Biophys. Acta* 192:479–85
13. Bernstein, L. H., Gutstein, S., Weiner, S., Efron, G. 1970. The absorption and malabsorption of folic acid and its polyglutamates. *Am. J. Med.* 48:570–79
14. Rosenberg, I. H., Godwin, H. A. 1971. The digestion and absorption of dietary folate. *Gastroenterology* 60:445–63
15. Reisenauer, A. M., Krumdieck, C. L., Halsted, C. H. 1977. Folate conjugase: two separate activities in human jejunum. *Science* 198:196–97
16. Hepner, G. W., Booth, C. C., Cowan, J., Hoffbrand, A. V., Mollin, D. L. 1968. Absorption of crystalline folic acid in man. *Lancet* 2:302–6
17. Gerson, C. D., Cohen, N., Hepner, G. W., Brown, N., Herbert, V., Janowitz, H. D. 1971. Folic acid absorption in man: enhancing effect of glucose. *Gastroenterology* 61:224–27
18. MacKenzie, J. F., Russell, R. I. 1976. The effect of pH on folic acid absorption in man. *Clin. Sci. Mol. Med.* 51:363–68
19. Blair, J. A., Matty, A. J. 1974. Acid microclimate in intestinal absorption. *Clin. Gastroenterol.* 3:183–95
20. Elsborg, L. 1974. Folic acid: a new approach to the mechanism of its intestinal absorption. *Dan. Med. Bull.* 21:1–11
21. Poe, M. 1977. Acidic dissociation constants of folic acid, dihydrofolic acid, and methotrexate. *J. Biol. Chem* 252:3724–28
22. Leslie, G. I., Rowe, P. B. 1972. Folate binding by the brush border membrane proteins of small intestinal epithelial cells. *Biochemistry* 11:1696–1703
23. Bhanthumnavin, K., Wright, J. R., Halsted, C. H. 1974. Intestinal transport of tritiated folic acid (^{3}H–PGA) in the everted gut sacs of different aged rats. *Johns Hopkins Med. J.* 135:152–60
24. Dhar, J. G., Selhub, J. J., Gay, C., Rosenberg, I. H. 1977. Characterization of the individual components of intestinal folate transport. *Gastroenterology* 72: 1049 (Abstr.)
25. Olinger, E. J., Bertino, J. R., Binder, H. J. 1973. Intestinal folate absorption II. Conversion and retention of pteroylmonoglutamate by jejunum. *J. Clin. Invest.* 52:2138–45
26. Selhub, J. H., Brin, H., Grossowicz, N. 1973. Uptake and reduction of radioactive folate by everted sacs of rat small intestine. *Eur. J. Biochem.* 33:433–38
27. Perry, J., Chanarin, I. 1973. Formylation of folate as step in physiological folate absorption. *Br. Med. J.* 1:558–59
28. Whitehead, V. M., Cooper, B. A., 1967. Absorption of unaltered folic acid from the gastrointestinal tract in man. *Br. J. Haematol.* 13:679–86
29. Luhby, A. L., Eagle, F. J., Roth, E., Cooperman, J. M. 1961. Relapsing megaloblastic anemia in an infant due to a specific defect in gastrointestinal absorption of folic acid. *Am. J. Dis. Child.* 102:482–83
30. Lanzkowski, P. 1970. Congenital malabsorption of folate. *Am. J. Med.* 48: 580–83
31. Lindenbaum, J., Whitehead, N., Reyner, F. 1975. Oral contraceptive hormones, folate metabolism, and the cervical epithelium. *Am. J. Clin. Nutr.* 28:346–53
32. Stephens, M. E. M., Craft, I., Peters, I. J., Hoffbrand, A. V. 1972. Oral contraceptives and folate metabolism. *Clin. Sci.* 42:405–14
33. Weir, D. G. 1974. The pathogenesis of folic acid deficiency in man. *Ir. J. Med. Sci.* 143:3–20
34. Klipstein, F. A. 1968. Tropical sprue. *Gastroenterology* 54:275–93

35. Swanson, V. L., Thomassen, R. W. 1965. Pathology of the jejunal mucosa in tropical sprue. *Am. J. Pathol.* 46:511–51
36. Hoffbrand, A. V., Douglas, A. P., Fry, L., Stewart, J. S. 1970. Malabsorption of dietary folate (pteroylpolyglutamate) in adult celiac disease and dermatitis herpetiformis. *Br. Med. J.* 4:85–89
37. Hoffbrand, A. V., Necheles, T. F., Maldonado, N., Horta, E., Santini, R. 1969. Malabsorption of folate polyglutamates in tropical sprue. *Br. Med. J.* 2:543–47
38. Gerson, C. D., Cohen, N., Brown, N., Lindenbaum, J., Hepner, G. W., Janowitz, H. D. 1974. Folic acid and hexose absorption in sprue. *Am J. Dig. Dis.* 19:911–19
39. Corcino, J. J. 1975. Recent advances in tropical sprue. In *Intestinal Absorption and Malabsorption,* ed. T. Z. Csaky, pp. 285–99. New York: Raven. 308 pp.
40. Corcino, J. J., Reisenauer, A. M., Halsted, C. H. 1976. Jejunal perfusion of simple and conjugated folates in tropical sprue. *J. Clin. Invest.* 58:298–305
41. Lucas, M. L., Cooper, B. T., Lei, F. H., Johnson, I. T., Holmes, G. K. T., Blair, J. A., Cooke, W. T. 1978. Acid microclimate in coeliac and Crohn's disease: a model for folate malabsorption. *Gut* 19:735–42
42. Halsted, C. H., Tamura, T. 1979. Folate deficiency in liver disease. In *Problems in Liver Diseases,* ed. C. S. Davidson, pp. 91–99. New York: Stratton Int. Med. Book Corp. 321 pp.
43. Halsted, C. H., Griggs, R. C., Harris, J. W. 1967. The effects of alcoholism on the absorption of folic acid (^{3}H–PGA) evaluated by plasma levels and urine excretion. *J. Lab. Clin. Med.* 69:116–31
44. Halsted, C. H., Robles, E. A., Mezey, E. 1971. Decreased jejunal uptake of folic acid (^{3}H–PGA) in alcoholic patients: roles of alcohol and nutrition. *N. Engl. J. Med.* 285:701–6
45. Halsted, C. H., Robles, E. A., Mezey, E. 1973. Intestinal malabsorption in folate deficient alcoholics. *Gastroenterology* 64:526–32
46. Russell, R. M., Dhar, G. J., Dutta, S. K., Rosenberg, I. H. 1979. Influence of intraluminal pH on folate absorption: studies in control subjects and in patients with pancreatic insufficiency. *J. Lab. Clin. Med.* 93:428–36
47. Reynolds, E. H. 1968. Mental effects of anticonvulsants, and folic acid metabolism. *Brain* 91:197–214
48. Dahlke, M. B., Mertens-Roesler, E. 1967. Malabsorption of folic acid due to diphenylhydantoin. *Blood* 30:341–51
49. Benn, A., Swan, C. H. J., Cooke, W. T., Blair, J. A., Matty, A. J., Smith, M. E. 1971. Effect of intraluminal pH on the absorption of pteroylmonoglutamic acid. *Br. Med. J.* 1:148–50
50. Perry, J., Chanarin, I. 1972. Observations on folate absorption with particular reference to folate polyglutamate and possible inhibitors to its absorption. *Gut* 13:544–50
51. Doe, W. F., Hoffbrand, A. V., Reed, P. I., Scott, J. M. 1971. Jejunal pH and folic acid. *Br. Med. J.* 1:669–70
52. Gerson, C. D., Hepner, G. W., Brown, N., Cohen, N., Herbert, V., Janowitz, H. D. 1972. Inhibition by diphenylhydantoin of folic acid absorption in man. *Gastroenterology* 63:246–51
53. Hoffbrand, A. V., Necheles, T. F. 1968. Mechanism of folate deficiency in patients receiving phenytoin. *Lancet* 2:528–30
54. Rosenberg, I. H., Godwin, H. A., Streiff, R. R., Castle, W. B. 1968. Impairment of intestinal deconjugation of dietary folate: a possible explanation of megaloblastic anemia associated with phenytoin therapy. *Lancet* 2:530–32
55. Baugh, C. M., Krumdieck, C. L. 1969. Effects of phenytoin on folic acid conjugase in man. *Lancet* 2:519–21
56. Franklin, J. L., Rosenberg, J. H. 1973. Impaired folic acid absorption in inflammatory bowel disease: effects of salicylazosulfapyridine (azulfidine). *Gastroenterology* 64:517–25
57. Selhub, J., Dhar, G. J., Rosenberg, I. H. 1978. Inhibition of folate enzymes by sulfasalazine. *J. Clin. Invest.* 61:221–24

Ann. Rev. Med. 1980. 31:89–96
Copyright © 1980 by Annual Reviews Inc. All rights reserved

PROSTAGLANDINS AND PLATELETS

◆7345

J. F. Mustard, M.D. and R. L. Kinlough-Rathbone, M.D.

Department of Pathology, McMaster University, Hamilton, Ontario, Canada

M. A. Packham, Ph.D.

Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada

INTRODUCTION

During formation of hemostatic plugs and thrombi, platelets (*a*) adhere to collagen and other components in the damaged vessel wall; (*b*) release granule contents including ADP; (*c*) aggregate; (*d*) accelerate the formation of thrombin; and (*e*) give rise to prostaglandins G_2 and H_2 (PGG_2 and PGH_2) and thromboxane A_2 (TXA_2). These compounds augment platelet aggregation and the release reaction. The damaged vessel wall also forms prostacyclin (PGI_2) that can inhibit platelet adherence, aggregation, and release. The enzymes involved in the formation of prostaglandins and TXA_2 in platelets and inhibitors of their formation are being actively investigated with the aim of inhibiting the contribution of platelets to thromboembolism (50, 51).

In tissues, the prostaglandins arise from 20-carbon polyunsaturated fatty acids in the phospholipids of cell membranes; arachidonic acid is the major precursor. However, the first prostaglandin tested for its effects on platelets was prostaglandin E_1 (PGE_1), which is formed from dihomo-γ-linolenic acid (30, 43, 47, 48). In 1966 Kloeze (42) demonstrated that PGE_1 inhibits platelet aggregation; other researchers subsequently found that it increases platelet cyclic AMP (31, 32). PGD_2 and PGI_2, which are formed from arachidonic acid, also increase cAMP in human platelets by stimulating adenylate cyclase. PGI_2 is much more potent than the other prostaglandins. Compounds such as dipyridamole that inhibit platelet phosphodiesterase

0066-4219/80/0401-0089$01.00

prevent the breakdown of cAMP and may augment the effects of these prostaglandins (32, 47). Increased cAMP may make internal calcium unavailable for the calcium-dependent platelet reactions that are involved in adherence, shape change, aggregation, and release (29–32).

A significant development in the 1970s was the demonstration by Samuelsson's group that platelets form PGG_2, PGH_2 and TXA_2 (58). These are short-lived compounds, and TXA_2 is a potent aggregating and release-inducing agent. Whether or not PGH_2 can itself induce aggregation without first being converted to TXA_2 is still an unsettled question (30).

ARACHIDONIC ACID CONVERSION

A major advance has been the delineation of the pathways involved in prostaglandin formation (Figure 1) (30, 44). A key step in the formation of prostaglandins is the freeing of arachidonate from membrane phospholipids.[1] When platelets are stimulated, phospholipase A_2 is activated, probably by mobilization of calcium, and catalyzes the cleavage of arachidonic acid

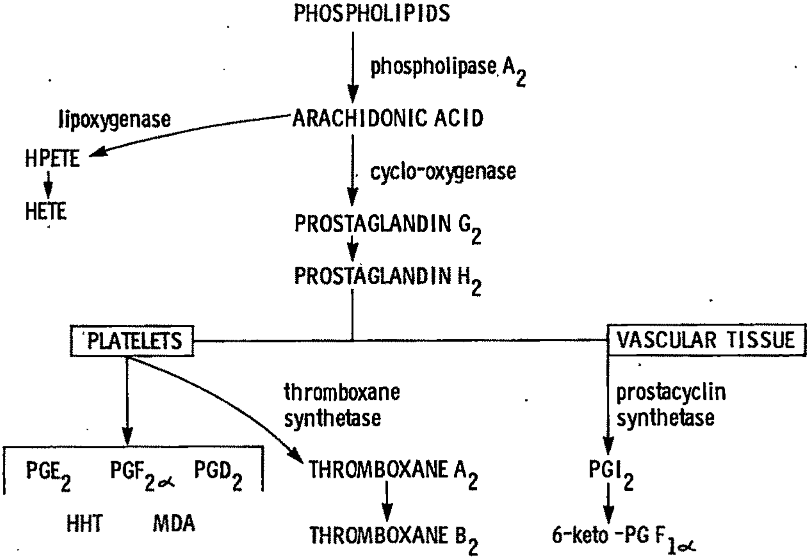


Figure 1 Metabolism of arachidonic acid and biosynthesis of prostaglandins (PG) and thromboxanes in platelets and vascular tissues (43). HPETE = 12-L-hydroperoxy – 5,8,10,14-eicosatetraenoic acid; HETE = 12-L-hydroxy – 5,8,10,14-eicosatetraenoic acid; HHT = 12-L-hydroxy –5,8,10-heptadecatrienoic acid; MDA = malondialdehyde.

---

[1] It has recently been recognized that phospholipase C and diglyceride lipase may be responsible for freeing arachidonate from membrane phospholipids instead of, or in addition to, phospholipase A_2. (Rittenhouse-Simmons, S. 1979. *J. Clin. Invest.* 63:580–87; Bell, R. L., Kennerly, D. A., Stanford, N., Majerus, P. W. 1979. *Proc. Natl. Acad. Sci. USA* 76:3238–41.)

from the second position on phospholipid molecules, principally phosphatidyl choline and phosphatidyl inositol (9, 10); this does not occur in the resting state. The release of platelet granule contents accompanies the activation of phospholipase A_2. Release-inducing agents, such as collagen and thrombin, have all been shown to activate phospholipase A_2.

If endothelial cells are mechanically stimulated or exposed to agents such as thrombin (20, 65, 66), their phospholipase A_2 is activated and they rapidly form PGI_2 from the arachidonic acid that is liberated (Figure 1). Stimulated smooth muscle cells form lesser amounts of PGI_2 (46).

INHIBITION OF THROMBOXANE A_2 FORMATION

Inhibition of Phospholipase A_2

Agents that inhibit the action of phospholipase A_2 or prevent its activation inhibit the removal of arachidonic acid from membrane phospholipids and thus block the formation of the products formed from it. These agents include steroidal anti-inflammatory drugs and their metabolites (28, 35), the antimalarial drug mepacrine (10), and other membrane-active drugs (25). In many cases, however, compounds that inhibit phospholipase A_2 also inhibit other enzymes so their effects are not specific. There is some uncertainty about the effect of the anti-inflammatory steroids on platelet phospholipase A_2 (59), although concentrations of methylprednisolone that inhibit platelet aggregation prevent phospholipase A_2 activation (J.-P. Cazenave and J. F. Mustard, unpublished observations). Since platelet aggregation and the release of granule contents can occur through mechanisms that do not involve activation of the arachidonate pathway (40), these drugs would be expected to cause only partial inhibition of platelet aggregation at a site of vessel wall injury.

Inhibition of Cyclo-Oxygenase

Aspirin inhibits platelet aggregation and the platelet release reaction because it causes irreversible acetylation of the cyclo-oxygenase of platelets (57). This inhibition persists for the life of the platelets and thus the effect of aspirin may be apparent as long as seven days after a single dose. By acetylating the cyclo-oxygenase, aspirin inhibits the conversion of freed arachidonate to the active intermediates PGG_2, PGH_2, and TXA_2 and the more stable end products. Cyclo-oxygenase is also inhibited by nonsteroidal anti-inflammatory drugs such as indomethacin, phenylbutazone, and fenoprofen and by sulfinpyrazone, but their effects are lost when they or their active metabolites are removed from the circulation (2, 13, 55).

However, inhibition of cyclo-oxygenase in vivo can not completely block platelet aggregation and the release of platelet granule contents because aggregation and release can occur through mechanisms independent of the products formed from arachidonic acid (40). Primary ADP-induced aggregation occurs without activation of the arachidonate pathway. The nonsteroidal anti-inflammatory drugs only partially inhibit thrombin-induced aggregation and release, because thrombin also can act through pathways independent of the arachidonate pathway (40).

Although the nonsteroidal anti-inflammatory drugs are weak or ineffective inhibitors of platelet adherence to collagen or of the release of granule contents from the platelets actually adherent to collagen (39), the drugs do inhibit the formation of TXA_2 and thereby inhibit collagen-induced platelet aggregation. However, experimental evidence that inhibition of the cyclo-oxygenase in platelets strongly inhibits thrombosis is far from conclusive and has only been obtained in some circumstances (51). For example, studies with arteriovenous shunts in dialysis patients have shown that aspirin or sulfinpyrazone, both of which inhibit cyclo-oxygenase, do diminish the frequency and extent of thrombus formation (15, 37).

Inhibition of Thromboxane Synthetase

There are more selective inhibitors of platelet aggregation and release due to products formed from arachidonate: they inhibit thromboxane synthetase without affecting other enzymes involved in the conversion of arachidonate to prostaglandins such as PGI_2 synthetase of the vessel wall. Thromboxane A_2 receptor antagonists have also been described (3, 30, 34, 60).

EFFECTS OF PGI_2 AND INHIBITION OF ITS FORMATION

It has been suggested that PGI_2, formed by vessel walls or in the lungs, is a circulating hormone and is responsible for preventing platelets from adhering to normal endothelium (47). However, even when PGI_2 formation is prevented by aspirin, platelets do not adhere to normal endothelium (22). When the endothelium is lost, platelets rapidly adhere to the exposed subendothelium; PGI_2 in high concentrations inhibits this adherence, but only by about 50% (17, 33, 64). Attempts to demonstrate that inhibition of PGI_2 production in vivo leads to increased platelet accumulation on the subendothelium of the aorta have failed (22), but the PGI_2 formed in large vessels where blood flow is rapid is unlikely to be present in sufficiently high local concentrations to influence platelet accumulation. Nevertheless, PGI_2 formed by the vessel wall may be an inhibitor of thrombosis associated with vessel injury under some circumstances.

PGE_1 and PGI_2 inhibit thrombus formation in extracorporeal circulations (1, 18, 47). In the microcirculation, where the volume of fluid in the vessels is small, stimulation of the vessel wall may trigger formation of sufficient PGI_2 to affect thrombosis, particularly if stasis occurs. Inhibition of PGI_2 formation in the microcirculation enhances platelet aggregate formation (12, 56). Since PGI_2 formation by the vessel wall, particularly when the wall is stimulated, may be important in limiting the extent of thrombus formation, concentrations of anti-inflammatory drugs that inhibit PGI_2 production may conceivably enhance thrombosis.

The platelet cyclo-oxygenase has been reported to be more sensitive to acetylation by aspirin than is the cyclo-oxygenase of the vessel wall (14). This observation has led to the suggestion that low doses of aspirin may be optimum for inhibition of thrombus formation, and that high doses may be less effective and even thrombogenic under some circumstances (16). There is some support for this concept from studies with experimental animals. Kelton and his colleagues (38) found that administration of high doses of aspirin, which inhibited PGI_2 formation, enhanced thrombosis produced by injury of the jugular veins of rabbits. In these studies thrombin probably stimulated formation of PGI_2, which in turn inhibited platelet aggregation and thrombosis. More extensive thrombosis occurred when sufficient aspirin was used to inhibit the increased PGI_2 formation at the injury site. Arfors' group (5, 8) also observed that doses of aspirin high enough to inhibit PGI_2 production did not inhibit thrombus formation at sites of vessel injury and caused more rapid arrest of bleeding than occurred in untreated animals. Inhibition of PGI_2 production in the jugular vein of thrombocytopenic rabbits allows the vessel wall to contract around a puncture site and shortens the bleeding time (11). Low doses of aspirin (0.3 g) prolong the bleeding time whereas in some circumstances high doses do not appear to have an effect on the bleeding time in man (54).

This experimental evidence indicates that high doses of aspirin can enhance thrombus formation in situations where PGI_2 formation would be strongly stimulated in the absence of aspirin; however, this may not occur in areas with less capacity to form PGI_2, such as on atherosclerotic plaques (21). There is no indication that the high doses of aspirin given to humans under some circumstances are thrombogenic. In contrast, in one study in which patients underwent knee surgery, high doses of aspirin were beneficial whereas low doses did not affect the incidence of thromboembolic complications (45). It is not known whether or not the high doses of aspirin were sufficient to inhibit PGI_2 formation.

Compounds that stimulate the vessel wall to produce PGI_2 could be antithrombotic; BAY g 6575 reportedly raises circulating PGI_2 levels in man (62). However, the hypotensive effect of such treatment (6, 27) is cause for concern.

CLINICAL TRIALS

In clinical trials, aspirin has had only a limited effect on those clinical complications of atherosclerosis considered to be thromboembolic (7, 19, 24, 26). If thromboembolism is the principal mechanism causing the clinical complications in patients with transient attacks of cerebral ischemia (TIA), aspirin might be expected to reduce the incidence and severity of the complications. In the Canadian Stroke Study (7) a beneficial effect of aspirin could only be shown if a number of endpoints (i.e. TIAs, strokes, and death) were used to compare the control and treated groups of patients; the benefit was restricted to males. If aspirin had a powerful effect on thrombosis and embolization, one would have expected the frequency of TIAs in this study to be greatly diminished but it was not.

In the Anturane Reinfarction Trial (4), sulfinpyrazone therapy, begun 25–35 days after myocardial infarction, reduced sudden cardiac deaths by 57%. This effect appeared to be restricted to the first six months after infarction. Sulfinpyrazone had no effect on subsequent myocardial infarction. In two studies, aspirin did not significantly reduce deaths after myocardial infarction (19, 24), but most of the patients did not receive aspirin until several months after infarction. However, in Elwood's study (24) one group of patients was treated with aspirin soon after infarction and showed a 40% reduction in cardiac deaths. Unfortunately, the study design did not permit conclusions to be drawn from these observations. To determine whether or not aspirin will have the same effect as sulfinpyrazone in the immediate post infarction period a controlled clinical trial will have to be carried out. Recent experiments in animals demonstrate that both aspirin and sulfinpyrazone can prevent sudden death associated with myocardial ischemia in circumstances where platelet aggregation does not occur (49, 63). The arachidonate pathway may be activated in ischemic tissue (36), giving rise to products that cause ventricular fibrillation. If so, aspirin and sulfinpyrazone would inhibit the formation of such products; an effect of this sort could overshadow their effects on platelets in the prevention of the clinical complications of arterial disease.

DIETARY FATTY ACIDS

The effect of dietary fatty acids on thromboembolic events and the development of atherosclerosis is attracting attention. The studies of dietary linoleic acid enrichment are inconclusive (52). Enriching diets with dihomo-γ-linolenic acid decreases the sensitivity of platelets to collagen and thrombin, probably because this fatty acid replaces arachidonic acid in the platelet phospholipids and gives rise to PGE_1, PGG_1, and TXA_1, instead of TXA_2

(48, 67). It was recently proposed that diets rich in eicosapentaenoic acid would decrease platelet sensitivity to aggregating agents since TXA_3 (which does not aggregate platelets) is formed from this fatty acid. In the vessel wall, PGI_3 is formed from eicosapentaenoic acid and is similar to PGI_2 in its ability to inhibit platelet reactions (23). These findings indicate that the thromboembolic complications of atherosclerosis may be modified by controlling the fatty acid content of the diet.

SUMMARY

Release of arachidonate from cell membrane phospholipids by activation of phospholipase A_2 is a key step in the formation of prostaglandins and thromboxanes. In platelets, PGH_2 and TXA_2 are formed from arachidonate and can cause aggregation and the release of granule contents; in vascular tissue, PGI_2 is formed instead and, by increasing platelet cAMP, inhibits platelet reactions. There is considerable interest in inhibitors of the enzymes in these pathways as drugs to modify thrombus formation. Results of the clinical trials, however, indicate that drugs which inhibit cyclo-oxygenase may not have a major effect on the thromboembolic complications of arterial disease.

Literature Cited

1. Addonizio, V. P. Jr., Strauss, J. F. III, Colman, R. W., Edmunds, L. H. Jr. 1979. *J. Thorac. Cardiovasc. Surg.* 77:119–26
2. Ali, M., McDonald, J. W. D. 1978. *Thromb. Res.* 13:1057–66
3. Allan, G., Eakins, K. E. 1978. *Prostaglandins* 15:659–61
4. Anturane Reinfarction Trial Research Group. 1978. *N. Engl. J. Med.* 298: 289–95
5. Arfors, K.-E., Bergqvist, D., Tangen, O. 1975. *Acta Chir. Scand.* 141:40–42
6. Armstrong, J. M., Lattimer, N., Moncada, S., Vane, J. R. 1978. *Br. J. Pharmacol.* 62:125–30
7. Barnett, H. J. M. et al. 1978. *N. Engl. J. Med.* 299:53–59
8. Bergqvist, D., Arfors, K.-E. 1976. *Haemostasis* 5:74–84
9. Bills, T. K., Smith, J. B., Silver, M. J. 1977. *J. Clin. Invest.* 60:1–6
10. Blackwell, G. J. 1978. *Adv. Prostaglandin Thromboxane Res.* 3:137–42
11. Blajchman, M. A., Senyi, A. F., Hirsh, J., Surya, Y., Buchanan, M. 1979. *J. Clin. Invest.* 63:1026–35
12. Bourgain, R. H. 1978. *Haemostasis* 7:252–55
13. Buchanan, M. R., Rosenfeld, J., Hirsh, J. 1978. *Thromb. Res.* 13:883–92
14. Burch, J. W., Baenziger, N. L., Stanford, N., Majerus, P. W. 1978. *Proc. Natl. Acad. Sci. USA* 75:5181–84
15. Burch, J. W., Harter, H., Stanford, N., Delmez, J. D., Majerus, P. W. 1979. *Clin. Res.* 27:509A
16. Burch, J. W., Stanford, N., Majerus, P. W. 1978. *J. Clin. Invest.* 61:314–19
17. Cazenave, J.-P., Dejana, E., Kinlough-Rathbone, R. L., Richardson, M., Packham, M. A., Mustard, J. F. 1979. *Thromb. Res.* 15:273–79
18. Coppe, D., Wonder, T., Snider, M., Salzman, E. 1978. *Circulation* 59:II-207
19. Coronary Drug Project Research Group. 1976. *J. Chronic Dis.* 29:625–42
20. Czervionke, R. L., Smith, J. B., Hoak, J. C., Fry G. L., Haycraft, D. L. 1979. *Thromb. Res.* 14:781–86
21. D'Angelo, V., Villa, S., Mysliwiec, M., Donati, M. B., de Gaetano, G. 1978. *Thromb. Haemostasis* 39:535–36
22. Dejana, E., Cazenave, J.-P., Groves, H. M., Kinlough-Rathbone, R. L., Mustard, J. F. 1979. *Fed. Proc.* 38:1271 (Abstr.)

23. Dyerberg, J., Bang, H. O., Stoffersen, E., Moncada, S., Vane, J. R. 1978. *Lancet* 2:117–19.
24. Elwood, P. C., Cochrane, A. L., Burr, M. L., Sweetnam, P. M., Williams, G., Welsby, E., Hughes, S. J., Renton, R. 1974. *Br. Med. J.* 1:436–40
25. Feinstein, M. B., Becker, E. L., Fraser, C. 1977. *Prostaglandins* 14:1075–93
26. Fields, W. S., Lemak, N. A., Frankowski, R. F., Hardy, R. J. 1977. *Stroke* 8:301–16
27. Fitzpatrick, T. M., Alter, I., Corey, E. J., Ramwell, P. W., Rose, J. C., Kot, P. A. 1978. *Circ. Res.* 42:192–94
28. Flower, R. 1978. *Adv. Prostaglandin Thromboxane Res.* 3:105–12
29. Gerrard, J. M., White, J. G. 1978. *Progr. Hemostasis Thromb.* 4:87–125
30. Gorman, R. R. 1979. *Fed. Proc.* 38:83–88
31. Harris, R. H., Ramwell, P. W. Gilmer, P. J. 1979. *Ann. Rev. Physiol.* 41:653–68
32. Haslam, R. J., Davidson, M. M. L., Davies, T., Lynham, J. A., McClenaghan, M. D. 1978. *Adv. Cyclic Nucleotide Res.* 9:533–52
33. Higgs, E. A., Moncada, S., Vane, J. R., Caen, J. P., Michel, H., Tobelem, G. 1978. *Prostaglandins* 16:17–22
34. Hokama, Y., Morita, A. H., Abad, M. A., Joyo, B. S., Uchida, D. A., Fischer, G. W. 1978. *Res. Commun. Chem. Pathol. Pharmacol.* 19:141–49
35. Hong, S.-C. L., Levine, L. 1976. *Proc. Natl. Acad. Sci. USA* 73:1730–34
36. Isakson, P. C., Raz, A., Hsueh, W., Needleman, P. 1978. Lipases and prostaglandin biosynthesis. *Adv. Prostaglandin Thromboxane Res.* 3:113–20
37. Kaegi, A., Pineo, G. F., Shimizu, A., Trivedi, H., Hirsh, J., Gent, M. 1974. *N. Engl. J. Med.* 290:304–6
38. Kelton, J. G., Hirsh, J., Carter, C. J., Buchanan, M. R. 1978. *J. Clin. Invest.* 62:892–95
39. Kinlough-Rathbone, R. L., Cazenave, J.-P., Packham, M. A., Mustard, J. F. 1979. *Lab. Invest.* In press
40. Kinlough-Rathbone, R. L., Packham, M. A., Reimers, H.-J., Cazenave, J.-P., Mustard, J. F. 1977. *J. Lab. Clin. Med.* 90:707–19
41. Deleted in proof
42. Kloeze, J. 1966. In *Prostaglandins,* Nobel Symp. 2, Stockholm, pp. 241–52, Stockholm: Almqvist & Wiksell
43. Lands, W. E. M. 1979. *Ann. Rev. Physiol.* 41:633–52
44. Malmsten, C. 1979. *Br. J. Haematol.* 41:453–58
45. McKenna, R., Bachmann, F., Galante, J., Kaushal, S. P., Meredith, P. 1976. *Blood* 48:977 (Abstr.)
46. Moncada, S., Herman, A. G., Higgs, E. A., Vane, J. R. 1977. *Thromb. Res.* 11:323–44
47. Moncada, S., Vane, J. R. 1979. *Fed. Proc.* 38:66–71
48. Moncada, S., Vane, J. R. 1979. *N. Engl. J. Med.* 300:1142–47
49. Moschos, C. B., Haider, B., De La Cruz, C. Jr., Lyons, M. M. Regan, T. J. 1978. *Circulation* 57:681–84
50. Mustard, J. F., Packham, M. A. 1978. In *Cardiovascular Drugs,* ed. G. S. Avery, 3:1–83. New York: Adis
51. Mustard, J. F., Packham, M. A. 1979. In *Controversies in Therapeutics,* ed. L. Lasagna. Philadelphia: Saunders
52. Mustard, J. F., Packham, M. A., Kinlough-Rathbone, R. L. 1978. In *Thrombosis: Animal and Clinical Models,* Vol. 102, *Adv. Exp. Biol. Med.,* ed. H. J. Day, B. A. Molony, E. E. Nishizawa, R. H. Rynbrandt, pp. 7–30. New York: Plenum
53. Deleted in proof
54. O'Grady, J., Moncada, S. 1978. *Lancet* 2:780
55. Rane, A., Oelz, O., Frolich, J. C., Seyberth, H. E., Sweetman, B. J., Watson, J. T., Wilkinson, G. R., Oates, J. A. 1978. *Clin. Pharmacol. Ther.* 23:658–68
56. Rosenblum, W. I., El-Sabban, F. 1978. *Circ. Res.* 43:238–41
57. Roth, G. J., Majerus, P. W. 1977. In *Prostaglandins in Hematology,* ed. M. J. Silver, J. B. Smith, J. J. Kocsis, pp. 345–60. New York: Spectrum
58. Samuelsson, B. 1977. See Ref. 57, pp. 1–10
59. Schoene, N. W. 1978. *Adv. Prostaglandin Thromboxane Res.* 3:121–26
60. Tai, H.-H., Yuan, B. 1978. *Biochem. Biophys. Res. Commun.* 80:236–42
61. Deleted in proof
62. Vermylen, J., Chamone, D. A. F., Verstraete, M. 1979. *Lancet* 1:518–20
63. Vik-Mo, H. 1978. *Scand. J. Haematol.* 21:225–32
64. Weiss, H. J., Turitto, V. T. 1979. *Blood* 53:244–50
65. Weksler, B. B., Ley, C. W., Jaffe, E. A. 1978. *J. Clin. Invest.* 62:923–30
66. Weksler, B. B., Marcus, A. J., Jaffe, E. A. 1977. *Proc. Natl. Acad. Sci. USA* 74:3922–26
67. Willis, A. L., Stone, K. J., Hart, M., Gibson, V., Marples, P., Botfield, E., Comai, K., Kuhn, D. C. 1977. See Ref. 57, 371–410

Ann. Rev. Med. 1980. 31:97–108
Copyright © 1980 by Annual Reviews Inc. All rights reserved

HIGH DENSITY LIPOPROTEINS AND ATHEROSCLEROSIS

❖7346

G. J. Miller, M.D., M.R.C.P.

UK Medical Research Council, Port-of-Spain, Trinidad, West Indies

The circulating lipoproteins range in diameter from less than 5 nm to greater than 500 nm. The larger the particle size the greater is its lipid content relative to protein and the lower is its density. These differences permit the separation of lipoproteins into density classes by ultracentrifugation (1). When suspended in a solution of density 1.063 g/ml and subjected to an intense centrifugal field, particles less dense than the solution float to the surface while the remainder sink as sediment. Many density classes can be prepared in this way (preparative ultracentrifugation) by suitable adjustment of the solvent density. The major classes currently defined are listed in Table 1.

Chylomicrons are not normally found in fasting plasma, and any present float to the surface after standing at 4°C for more than 24 hours. In the older nomenclature still in use, the term low density lipoprotein covered the range 1.006–1.063 g/ml. Total high density lipoprotein (HDL) is considered to be $HDL_2 + HDL_3$.

Several methods exist for the quantitation of the separated lipoproteins. In the earliest procedure (analytical ultracentrifugation), the lipoproteins are identified by flotation rate (units being the Svedburg of flotation) and quantified according to the area occupied within the photographic record obtained with a schlieren optical system. Lipoprotein concentrations are described more easily, however, by lipid analysis (2) as lipoprotein cholesterol or triglyceride.

While ultracentrifugation remains the reference method, simpler techniques of separation give similar results. These include precipitation, immunoassay, and electrophoresis. The separation of total HDL is usually achieved by precipitation of LDL, IDL, and VLDL with polyanions and divalent cations (3), especially heparin and manganese (4), leaving HDL in

0066-4219/80/0401-0097$01.00

Table 1 Density classes of circulating lipoproteins

Lipoprotein class	Density range (g/ml)
Very low density lipoprotein (VLDL) and chylomicrons	Less than 1.006
Intermediate density lipoprotein (IDL)	1.006–1.019
Low density lipoprotein (LDL)	1.019–1.063
High density lipoprotein 2 (HDL_2)	1.063–1.125
High density lipoprotein 3 (HDL_3)	1.125–1.210

the supernatant. Care must be taken to ensure complete precipitation of LDL and to obtain an accurate estimate of cholesterol concentration. Small differences in HDL cholesterol concentration are associated with large differences in liability to coronary heart disease (CHD), so that measurement error can lead to faulty prognosis. Storage of whole plasma at –20°C before separation of the lipoproteins, especially for more than 3 months, can lead to an underestimate of HDL cholesterol concentration. About 5% of total HDL as separated by preparative ultracentrifugation consists of an HDL species that is heparin-manganese precipitable (HDL_1) (5). Thus the two methods measure very similar but not identical HDL fractions. Immunodiffusion assays against the peptides apoprotein A-I and A-II (which are, except in rare instances, confined to HDL in fasting plasma) permit quantitation of total HDL (including HDL_1 apo A-I) in terms of its protein content (6). Electrophoretic techniques, which rely on differences in electrical charge between lipoproteins, are convenient but not easy to quantify accurately.

The only obligatory characteristic of HDL particles is that they possess a density between 1.063 and 1.210 g/ml; in fact they are a very heterogeneous group (7). Circulating particles consist normally of a core of apolar cholesterol ester and triglyceride enveloped by a shell of specific apoproteins and phospholipids. The hydrophobic fatty acid chains of the surface phospholipids are orientated toward the core. Free cholesterol appears to be at the core-shell interface. The HDL_3 subclass has a smaller mean particle diameter (5 nm) and lipid-to-protein ratio (0.8) than HDL_2 (8 nm and 1.4 respectively). The lipid composition of both subclasses is about 38% phosphatidylcholine, 27% cholesterol ester and 6% triglyceride by weight; other phospholipids, glycerides, and free cholesterol account for the remainder.

High density lipoproteins contain at least eight recognized apoproteins (7). Apoprotein A-I and A-II account for about 66% and 25% of total HDL protein respectively and the remainder comprises the minor peptides apo C-I, C-II, C-III, D, E, and a trace of B (the major apoprotein of LDL). Those particles containing E and B reside in HDL_2 and are precipitated by the heparin-manganese procedure.

PLASMA HDL AND CORONARY HEART DISEASE

The earliest lipoprotein studies in CHD revealed an array of abnormalities that could be termed an atheromatous dyslipoproteinemia. In its simplest form, it was a reduction in HDL total particle concentration and an increase in the particle concentration of lipoproteins of density less than 1.063 g/ml (8). Because high plasma lipid levels were believed responsible for atherosclerosis, attention was focused on those lipoproteins that were increased in concentration. High density lipoproteins were considered relatively unimportant since they made no direct contribution to increased lipid concentrations. Researchers looked for the atherogenic lipoprotein or combination of lipoproteins within the LDL, IDL, VLDL spectrum. Repeated reports of a low HDL concentration in atherosclerosis (9, 10) failed to create interest, until the "HDL hypothesis" proposed that HDL cleared cholesterol from extrahepatic tissues (11). During turnover, cholesterol must be transported within the circulating lipoproteins back to the liver, the only organ which can catabolize and excrete cholesterol in significant quantity. Should this pathway require HDL, and should HDL formation be rate limiting for this section of the cholesterol transport cycle, then a low HDL particle concentration might be associated with impaired clearance of cholesterol from the arterial intima.

The relation between HDL and CHD has been examined cross-sectionally in nearly 7000 men and women over 39 years old from five populations in the US (12). In every community the mean HDL cholesterol concentration was lower in subjects with CHD than in the remainder. The difference was often small (3–6 mg/100 ml) but statistically significant and independent of other recognized lipid risk factors. Furthermore, the relation applied within each age group and for each category of CHD. In men 50–69 years old, CHD prevalence (per 1000 subjects) ranged from 180 in those with an HDL cholesterol of less than 24 mg/100 ml, to 78 in those with an HDL cholesterol of more than 45 mg/100 ml.

Recent results from prospective studies confirm that a low HDL concentration precedes clinical CHD and that the association is independent of other coronary risk factors. In the Norwegian municipality of Tromsø, 6595 men 20–49 years old were observed, beginning in 1974 (13). By 1976, 21 men had had a coronary event. Results available for 17 of these 21 men were compared with those of 31 controls matched for age, area of residence, and reported habitual physical activity. Mean plasma HDL cholesterol concentrations were 26 and 39 mg/100 ml in cases and controls respectively ($P < 0.005$), while LDL cholesterol averaged 222 and 190 mg/100 ml respectively ($P < 0.005$). HDL cholesterol was three times more effective than LDL cholesterol in distinguishing between cases and controls. Plasma HDL cholesterol concentrations reported were generally low, probably owing to storage of the samples as whole plasma for two years before analysis, but this would not have affected the validity of internal comparisons. Mean apoprotein A-I concentrations were 105 and 140 mg/100 ml in cases and controls respectively ($P < 0.02$) (14). A similar study has been reported from Oslo (15) where nearly 15,000 men 40–49 years old were examined in 1972–1973. Men who developed CHD by 1977 had had as a group a 10% lower HDL cholesterol concentration at entry to the study than others still free of overt disease. This significant difference remained after allowance for triglyceride concentration and smoking habit.

In Framingham, Massachusetts, nearly 2500 men and women 49–82 years old were observed for about four years (16). Here also HDL cholesterol concentration emerged as the most potent lipid risk factor in both sexes ($P < 0.001$). The association applied to all forms of CHD and was independent of other coronary risk factors. Plasma LDL cholesterol concentration was correlated positively but less powerfully with the incidence of CHD ($P < 0.05$). In Israel approximately 10,000 men 40 years or older have been followed since 1963. Serum HDL cholesterol concentration was measured sometime later in an apparently representative subsample of 6562 men (17). A significant inverse association between HDL cholesterol and subsequent CHD, independent of other risk factors, was present in men over the age of 45 years, but not in men younger than 45 years.

The great majority of comparisons of lipoprotein concentrations in myocardial infarction survivors and subjects without CHD have shown low HDL concentrations in the survivors (18, 19). Albers et al (18) found a significant reduction in the ratio of HDL cholesterol to apoprotein A-I in those with CHD, perhaps owing to a disproportionately large reduction in HDL_2 compared to that in HDL_3 (9). Further support is provided by coronary angiography. In one study (20), the lowest HDL cholesterol concentrations were found in patients with extensive coronary disease while in another (21) the lowest concentrations were associated with left main coronary disease.

HDL METABOLISM IN RELATION TO ATHEROSCLEROSIS

Human metabolism of HDL is beginning to be understood, and there is sufficient information with which to construct a working theory for further study. Unlike chylomicrons and VLDL, HDL particles are not secreted as such into the circulation but appear to be assembled within the plasma from precursor components derived from the small intestine, liver, and, in the case of HDL cholesterol, the extrahepatic tissues and other circulating lipoproteins.

The intestine synthesizes apoprotein A-I and probably also A-II as surface components of mesenteric chylomicrons (22). The triglyceride core of these particles is hydrolyzed by extrahepatic lipoprotein lipase, and the fatty acids released are taken up by the tissues or transferred to albumin. Lipoprotein lipase is activated by apoprotein C-II present in HDL, VLDL, and chylomicrons. During lipolysis, apoprotein A-I, apo-C peptides and phospholipid leave the surface shell and appear within the HDL density range. Apoprotein A-I has an avidity for phospholipid, with which it forms disc-like particles in vitro. These and other observations led Tall & Small (23) to propose that with shrinkage of the chylomicron core during lipolysis, the surface monolayer of the particle is thrown into bilayered folds that break away as discrete particles. These apoprotein-phospholipid discs appear to be a nascent form of HDL.

Extrahepatic lipoprotein lipase is located on the capillary endothelial cell surface of such tissues as adipose tissue, muscle, and heart. Its activity fluctuates according to local requirements for triglyceride. For example, muscle lipoprotein lipase activity increases with high levels of physical activity (24). Lipolysis of chylomicrons by this enzyme would not only release fatty acid for energy purposes but also perhaps deliver in high local concentration to the same tissues nascent HDL particles rich in apoprotein A-I. Lipolysis of hepatic VLDL would not release an identical nascent HDL since the surface monolayer of these particles does not possess apoprotein A-I (7).

Lipoproteins circulate in association with the enzyme lecithin-cholesterol acyltransferase (LCAT), which catalizes the esterification of free cholesterol with the formation of lysolecithin from lecithin. Apoprotein A-I (25) and perhaps apo C-I and apo D act as cofactors; nascent HDL appears to be the preferred substrate (26). Thus LCAT may esterify nascent HDL free cholesterol and the ester product may then migrate into the HDL particle interior. The surface monolayer will then be free to acquire more unesterified cholesterol from circulating lipoproteins and cell membranes.

Circulating HDL appears to acquire cell cholesterol by binding to the surface membrane, apparently in proportion to local particle concentration. Very little HDL actually enters the cell (31). This binding interferes for a short period with LDL uptake by the cell, but thereafter cell cholesterol content decreases and LDL surface binding and uptake increase (32). A sequence of studies has demonstrated the removal of cholesterol from cells, first by whole serum (27) then by serum HDL (28) and certain HDL peptide-phospholipid recombinant particles (29). Nestel & Miller (30) found that cholesterol lost from adipocytes during weight reduction moves only into HDL and not into other circulating lipoproteins. This might explain the inverse relation between body cholesterol content and HDL cholesterol concentration (11).

As nascent HDL formation, cell binding, and cholesterol acquisition proceed, the particles become increasingly spherical, enlarged, less dense, and less suitable as substrate for LCAT (23). Concurrently, the chylomicrons decrease in size, increase in density, and become less suitable as substrate for lipoprotein lipase. Both reactions are probably integrated with cellular requirements for cholesterol and triglyceride. Should nascent HDL behave in this manner, then the first circulating product might be HDL_3 since this is intermediate in size between nascent HDL and HDL_2. Subsequent phases might involve VLDL of hepatic origin, for when these particles are incubated with HDL_3 and lipoprotein lipase, surface components and cholesterol are transferred from VLDL to HDL_3. The latter are thereby converted to particles that resemble HDL_2 (33). These interrelations may explain why the increase in lipoprotein lipase activity after physical training (24) appears to be associated with an increase in HDL_2 concentration (34) and reduction in VLDL concentration (35).

Rat liver secretes a nascent HDL rich in apoprotein E (36). Whether similar particles are secreted by the human liver is uncertain, although nascent HDL is present in human hepatic venous blood (37). HDL particles containing apo E are present in human blood, particularly in response to a high cholesterol diet (38), and can compete powerfully with LDL for high-affinity binding sites on the peripheral cell surface membrane (5). Other HDL subclasses bind to separate receptors on the cell surface (39). The physiological role of HDL containing apo E remains to be determined.

The fate of circulating HDL is uncertain, although some studies suggest that apoprotein A-I catabolism occurs in the hepatic parenchyma (40). Other studies suggest that HDL free cholesterol is the preferred source of cholesterol for biliary excretion and conversion to bile acid in man (41). These observations, and the low rate of uptake of HDL by some extrahepatic tissue cells in culture (31), suggest that the human liver might be an important site of HDL catabolism.

A lipoprotein designed to scavenge for cholesterol should be able to move freely in the extracellular space without being trapped in the mesh of glycosaminoglycans and without being engulfed by tissue cells. Circulating HDL appears to possess such qualities, for unlike LDL and VLDL these particles do not complex readily with heparin (a glycosaminoglycan) in vitro (3). These characteristics might explain why HDL makes little if any contribution to the lipid of atheromatous lesions.

HDL IN PERSPECTIVE

The HDL hypothesis offers an explanation for CHD in patients without hyperlipidemia. It also helps explain the relative protection of women as compared with men from CHD since in nearly all communities studied women have had relatively high HDL concentrations (11). More importantly, the hypothesis redresses the balance of cardiovascular research, which previously focused only on LDL, IDL, and VLDL. Plasma HDL must be integrated into a scheme that describes the relations of all circulating lipoproteins to CHD. Low density lipoproteins remain in this scheme since they are undoubtedly the direct source of most lipid in atheroma. So also must VLDL be considered, since its concentration is linked closely with HDL, especially HDL_2 (42), probably through the lipoprotein lipase system. Furthermore, VLDL appears to be the major source of LDL in humans (43).

Across a range of communities, mean HDL cholesterol concentration decreases as mean LDL and VLDL concentrations increase. This reciprocal pattern seems to be pathological, for CHD incidence increases as it becomes more marked (12, 13, 16). Against this, data from communities in which CHD incidence is very low suggest that a low HDL concentration is physiological in the presence of a low LDL concentration (44), and that a high LDL concentration might be physiological in the presence of a high HDL concentration (45). Natural estrogens increase both VLDL and HDL concentrations (46), while androgens exert the reverse effect (47). Thus the physiological correlations appear to be positive. These observations suggest that ratios such as LDL:HDL might provide an improved estimate of CHD risk, but this remains to be established.

Relations between the plasma lipoproteins and other atheromatous diseases are not so clearly demonstrable as are those with CHD (48). Several studies report a low mean HDL cholesterol concentration in cerebrovascular disease (49) and low HDL cholesterol and apoprotein A-I concentrations in peripheral vascular disease (50–52), but further study is required of HDL in these conditions.

HDL AND PREVENTION OF CORONARY HEART DISEASE

Estimates of CHD risk rest on an assessment of all lipoproteins. The mistake in the past was to ignore HDL; to ignore LDL and VLDL now would be similarly mistaken. In theory, HDL and LDL concentrations might be correlated by their roles in cell cholesterol metabolism. Prolonged exposure of cultured fibroblasts to HDL increases the cellular uptake of LDL (32) and presumably a suboptimal exposure to HDL would lead to repletion of the cell with cholesterol and suppression of LDL uptake. The importance of the splanchnic, hepatic, intravascular, and peripheral components of the lipid transport system for this ratio, and for the atherosclerotic process, require definition.

Dietary Polyunsaturated Fat and HDL

Dietary polyunsaturated fat reduces LDL concentration in many subjects, but its effects on HDL metabolism are less certain. Limited evidence from human study suggests that the acute, short-term, and long-term effects of dietary polyunsaturated fat on HDL might differ significantly. If 40% of a diet's calories are fat, a high ratio of polyunsaturated to saturated fat (P : S) has no significant effect on HDL after two weeks (53). Given for two weeks to two months, diets with a P : S ratio of between 1.0 and 4.0 appeared to reduce plasma apo A-I and sometimes also HDL cholesterol concentration, especially HDL_2 (54). Long-term studies are not available for normolipidemic subjects. In hypercholesterolemic men, some of whom were also hypertriglyceridemic, adherence for four years to a 30% fat diet with an estimated P:S ratio of 1.0 was associated with an increase in HDL cholesterol and a reduction in LDL cholesterol concentration (55). In another study, good adherence to this type of regime for one year was associated with an increase in mean HDL cholesterol concentration from 48 to 52 mg/100 ml (56). In a comparison of men in Stockholm and Edinburgh, those in Stockholm had the higher content of polyunsaturated fat in the plasma and adipose tissue (and presumably therefore in the diet) together with the higher HDL cholesterol concentration (57). Mean LDL cholesterol concentration did not differ significantly between the two communities.

Other Dietary Factors

Plasma HDL concentration is sensitive to acute changes in the proportion of calories supplied as dietary fat and carbohydrate. Change to a diet containing 80% of calories as carbohydrate and less than 1% as fat reduced HDL cholesterol and apoprotein A concentrations in five days (58). Long-term effects of carbohydrate-rich and fat-rich diets are less certain, although in one study HDL cholesterol concentration appeared to be independent of the proportion of calories consumed as fat over the year (59).

An acute change from a cholesterol-free diet to one that contained 1000 mg of cholesterol per day was accompanied after 2–4 weeks by an increase in HDL (and LDL) cholesterol concentration (60, 61), probably owing in part to an increase in the apo E subclass (38). Vegetarians have relatively low HDL and LDL concentrations and a relatively low LDL:HDL ratio (62). Addition of 16 g of crude fiber to the daily diet for four weeks did not effect HDL concentration (60). A daily supplement of 50 g of unprocessed wheat bran to the diet of five patients with type IV hyperlipoproteinemia was followed by an increase in HDL cholesterol concentration at three months which thereafter receded toward initial levels (63).

Other Factors of Importance for HDL Concentration

Any study of diet, lipoproteins, and atherosclerosis must control for other factors that affect the plasma lipoproteins. Drugs such as alcohol (64), phenytoin (65), and estrogen (46) increase HDL cholesterol concentration while others such as propranolol reduce the concentration (66). Intensive physical training is associated with an increase in HDL, mostly as HDL_2 (34, 35). Loss of weight is sometimes (67) although apparently not always (68) followed by an increase in HDL concentration, and HDL cholesterol concentration has been observed to decrease during acute caloric restriction with weight loss (69). Cigarette consumption is inversely related to HDL cholesterol concentration (70). Febrile illness can induce a reduction in HDL cholesterol concentration that persists for several weeks (71).

CONCLUSION

The functions of HDL particles are uncertain, but they might regulate the cholesterol content of extrahepatic tissues. The majority appear to be products of a system that promotes the efflux of cholesterol from peripheral cells and other circulating lipoproteins, presumably for transport to the liver for catabolism and excretion. A minor species that contains apoprotein E might moderate entry of LDL into peripheral cells by competing for specific receptors on the cell surface membrane. Convincing evidence for a low plasma HDL concentration in CHD emphasizes the need for a better understanding of HDL metabolism, its relations with other lipoproteins, and the effects of such factors as diet, energy expenditure, cigarette smoking, and alcohol consumption. The importance of HDL for cerebral and peripheral atherosclerosis needs careful study. Clinical judgements concerning CHD require assessment of the complete lipoprotein spectrum including HDL.

Literature Cited

1. Gofman, J. W., DeLalla, O., Glazier, F., Freeman, N. K., Lindgren, F. T., Nichols, A. V., Strisower, B., Tamplin, A. R. 1954. The serum lipoprotein transport system in health, metabolic disorders, atherosclerosis and coronary heart disease. *Plasma* 2:413–84
2. Havel, R. J., Eder, H. A., Bragdon, J. H. 1955. The distribution and chemical composition of ultracentrifugally separated lipoproteins in human serum. *J. Clin. Invest.* 34:1345–53
3. Burstein, M., Scholnick, H. R. 1973. Lipoprotein-polyanion-metal interactions. *Adv. Lipid Res.* 11:67–108
4. Warnick, G. R., Albers, J. J. 1978. A comprehensive evaluation of the heparin-manganese precipitation procedure for estimating high density lipoprotein cholesterol. *J. Lipid Res.* 19:65–76
5. Innerarity, T. L., Mahley, R. W., Weisgraber, K. H., Bersot, T. P. 1978. Apoprotein (E-A-II) complex of human plasma lipoproteins. II. Receptor binding activity of a high density lipoprotein subfraction modulated by the apo (E-A-II) complex. *J. Biol. Chem.* 253:6289–95
6. Cheung, M. C., Albers, J. J. 1977. The measurement of apolipoprotein A-I and A-II levels in men and women by immunoassay. *J. Clin. Invest.* 60:43–50
7. Smith, L. C., Pownall, H. J., Gotto, A. M. 1978. The plasma lipoproteins:

structure and metabolism. *Ann. Rev. Biochem.* 47:751–77

8. Barr, D. P., Russ, E. M., Eder, H. A. 1951. Protein-lipid relationships in human plasma. II. In atherosclerosis and related conditions. *Am. J. Med.* 11: 480–93
9. Gofman, J. W., Young, W., Tandy, R. 1966. Ischemic heart disease, atherosclerosis and longevity. *Circulation* 34: 679–97
10. Brunner, D., Löbl, K. 1958. Serum cholesterol, electrophoretic lipid pattern, diet and coronary heart disease: a study in coronary patients and in healthy men of different origin and occupation in Israel. *Ann. Intern. Med.* 49:732–50
11. Miller, G. J., Miller, N. E. 1975. Plasma-high-density lipoprotein concentration and development of ischaemic heart disease. *Lancet* 1:16–19
12. Castelli, W. P., Doyle, J. T., Gordon, T., Hames, C. G., Hjortland, M. C., Hulley, S. B., Kagan, A., Zukel, W. J. 1977. HDL cholesterol and other lipids in coronary heart disease. The Co-operative Lipoprotein Phenotyping Study. *Circulation* 55:767–72
13. Miller, N. E., Førde, O. H., Thelle, D. S., Mjøs, O. D. 1977. The Tromsø Heart-Study. High-density lipoprotein and coronary heart-disease: a prospective case-control study. *Lancet,* 1:985–8
14. Ishikawa, T., Fidge, N., Thelle, D. S., Førde, O. R., Miller, N. E. 1978. The Tromsø Heart Study: Serum apolipoprotein AI concentration in relation to future coronary heart disease. *Eur. J. Clin. Invest.* 8:179–82
15. Enger, S. C., Hjermann, I., Foss, O. P., Helgeland, A., Holme, I., Leren, P., Norum, K. R. 1979. High density lipoprotein cholesterol and myocardial infarction or sudden coronary death: a prospective case-control study in middle-aged men of the Oslo Study. *Artery.* In press
16. Gordon, T., Castelli, W. P., Hjortland, M. C., Kannel, W. B., Dawber, T. R. 1977. High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study. *Am. J. Med.* 62:707–14
17. Goldbourt, U., Medalie, J. H. 1979. High density lipoprotein cholesterol and incidence of coronary heart disease —The Israeli Ischaemic Heart Disease Study. *Am. J. Epidemiol.* 109:296–306
18. Albers, J. J., Cheung, M. C., Hazzard, W. R. 1978. High-density lipoproteins in myocardial infarction survivors. *Metabolism* 27:479–85
19. Avogaro, P., Bittolo Bon, G., Cazzolato, G., Quinci, G. B. 1979. Are apolipoproteins better discriminators than lipids for atherosclerosis? *Lancet* 1:901–3
20. Jenkins, P. J., Harper, R. W., Nestel, P. J. 1978. Severity of coronary atherosclerosis related to lipoprotein concentration. *Br. Med. J.* 2:388–91
21. Pearson, T. A., Bulkley, B. H., Achuff, S. C., Kwiterovitch, P. O., Gordis, L. 1978. The association of low levels of HDL cholesterol and left main coronary artery disease. *Am. J. Epidemiol.* 108:241
22. Glickman, R. M., Green, P. H. R., Lees, R. S., Tall, A. R. 1978. Apoprotein A-I synthesis in normal intestinal mucosa and in Tangier disease. *N. Engl. J. Med.* 299:1424–27
23. Tall, A. R., Small, D. M. 1978. Plasma-high density lipoproteins. *N. Engl. J. Med.* 299:1232–36
24. Nikkilä, E. A., Taskinen, M.-R., Rehunen, S., Härkönen, M. 1978. Lipoprotein lipase activity in adipose tissue and skeletal muscle of runners: relation to serum lipoproteins. *Metabolism* 27: 1661–71
25. Fielding, C. J., Shore, V. G., Fielding, P. E. 1972. A protein co-factor of lecithin-cholesterol acyltransferase. *Biochem. Biophys. Res. Commun.* 46:1493–98
26. Hamilton, R. L., Williams, M. C., Fielding, C. J., Havel, R. J. 1978. Discoidal bilayer structure of nascent high density lipoproteins from perfused rat liver. *J. Clin. Invest.* 58:667–80
27. Rutenberg, H. L., Soloff, L. A. 1971. Possible mechanism of egress of free cholesterol from the arterial wall. *Nature* 230:123–25
28. Bondjers, G., Björkerud, S. 1975. HDL-dependent elimination of cholesterol from human arterial tissue. *Proc. Eur. Soc. Clin. Invest.* 9:51–52
29. Stein, O., Vanderhoek, J., Stein, Y. 1977. Cholesterol ester accumulation in cultured aortic smooth muscle cells. Induction of cholesterol ester retention by chloroquine and low density lipoprotein and its reversion by mixtures of high-density apolipoprotein and sphingomyelin. *Atherosclerosis* 26:465–82
30. Nestel, P. J., Miller, N. E. 1978. Mobilization of adipose tissue cholesterol in high density lipoprotein during weight reduction in man. In *High Density Lipoproteins and Atherosclerosis,* ed. A. M. Gotto Jr., N. E. Miller, M. F. Oliver, pp 51–54. Amsterdam/New York: Elsevier/North-Holland, 227 pp.

31. Miller, N. E., Weinstein, D. B., Steinberg, D. 1977. Binding, internalization, and degradation of high density lipoprotein by cultured human fibroblasts. *J. Lipid Res.* 18:438–50
32. Miller, N. E. 1978. Induction of low density lipoprotein receptor synthesis by high density lipoprotein in cultures of human skin fibroblasts. *Biochim. Biophys. Acta* 529:131–37
33. Patsch, J. R., Gotto, A. M., Olivecrona, T., Eisenberg, S. 1978. Formation of high density lipoprotein$_2$-like particles during lipolysis of very low density lipoproteins *in vitro*. *Proc. Natl. Acad. Sci. USA* 75:4519–23
34. Krauss, R. M., Lindgren, F. T., Wood, P. D., Haskell, W. L., Albers, J. J., Cheung, M. C. 1977. Differential increases in plasma high density lipoprotein subfractions and apolipoproteins (apo-LP) in runners. *Circulation* 56:Suppl. 3, p. 4
35. Lehtonen, A., Viikari, J. 1978. Serum triglycerides and cholesterol and serum-high density lipoprotein cholesterol in highly physically active men. *Acta Med. Scand.* 204:111–14
36. Felker, T. E., Fainaru, M., Hamilton, R. L., Havel, R. J. 1977. Secretion of the arginine-rich and A-I apolipoproteins by the isolated perfused rat liver. *J. Lipid Res.* 18:465–73
37. Turner, P., Miller, N., Chrystie, I., Coltart, J., Mistry, P., Nicoll, A., Lewis, B. 1979. Splanchnic production of discoidal plasma high-density lipoprotein in man. *Lancet* 1:645–47
38. Mahley, R. W., Innerarity, T. L., Bersot, T. P., Lipson, A., Margolis, S. 1978. Alterations in human high-density lipoproteins, with or without increased plasma cholesterol, induced by diets high in cholesterol. *Lancet* 2:807–9
39. Koschinsky, T., Carew, T. E., Steinberg, D. 1977. A comparative study of surface binding of human low density and high density lipoproteins to human fibroblasts: regulation by sterols and susceptibility to proteolytic digestion. *J. Lipid Res.* 18:451–58
40. Van Berkel, T. J. C., Koster, J. F., Hülsmann, W. C. 1977. High density lipoprotein and low density lipoprotein catabolism by human liver and parenchymal and non-parenchymal cells from rat liver. *Biochim. Biophys. Acta* 486:586–89
41. Halloran, L. G., Schwartz, C. C., Vlahcevic, Z. R., Hisman, R. M., Swell, L. 1978. Evidence for high-density lipoprotein-free cholesterol as the primary precursor for bile-acid synthesis in man. *Surgery* 84:1–7
42. Nichols, A. V. 1967. Human serum lipoproteins and their interrelationships. *Adv. Biol. Med. Phys.* 11:109–58
43. Sigurdsson, G., Nicoll, A., Lewis, B. 1975. Conversion of very low density lipoprotein to low density lipoprotein. *J. Clin. Invest.* 56:1481–90
44. Connor, W. E., Cerqueira, M. T., Connor, R. W., Wallace, R. B., Malinow, M. R., Casdorph, H. R. 1978. The plasma lipids, lipoproteins, and diet of the Tarahumara Indians of Mexico. *Am. J. Clin. Nutr.* 31:1131–42
45. Bang, H. O., Dyerberg, J., Nielsen, A. B. 1971. Plasma lipid and lipoprotein pattern in Greenlandic West-Coast Eskimos. *Lancet* 1:1143–46
46. Krauss, R. M., Lindgren, F. T., Wingerd, J., Bradley, D. D., Ramcharan, S. 1979. Effects of estrogens and progestins on high density lipoproteins. *Lipids* 14:113–18
47. Solyom, A. 1972. Effect of androgens on serum lipids and lipoproteins. *Lipids* 7:100–5
48. Kannel, W. B., Gordon, T., Dawber, T. 1974. Role of lipids in the development of brain infarction—The Framingham Study. *Stroke* 5:679–85
49. Rössner, S., Kjellin, K. G., Mettinger, K. L., Sidén, A., Söderström, C. E. 1978. Dyslipoproteinemia in patients with ischemic cerebro-vascular disease. A study of stroke before the age of 55. *Atherosclerosis* 30:199–209
50. Carlson, L. A. 1973. Plasma lipids and atherosclerosis. *J. Clin. Pathol.* 26:Suppl. 5, pp. 43–47
51. Zilcher, H., Kaliman, J., Muller, M. 1979. H.D.L. cholesterol in peripheral vascular disease. *Lancet* 1:558–59
52. Vergani, C., Trovato, G., Dioguardi, N. 1978. Serum total lipids, liproprotein cholesterol, apoproteins A and B in cardiovascular disease. *Clin. Chim. Acta* 87:127–33
53. Chait, A., Onitiri, A., Nicoll, A., Rabaya, E., Davies, J., Lewis, B. 1974. Reduction of serum triglyceride levels by polyunsaturated fat. *Atherosclerosis* 20:347–64.
54. Shepherd, J., Packard, C. J., Patsch, J. R., Gotto, A. M., Taunton, O. D. 1978. Effects of dietary polyunsaturated and saturated fat on the properties of high density lipoproteins and the metabolism of apolipoprotein A-I. *J. Clin. Invest.* 61:1582–92
55. Hjermann, I., Enger, S. C., Helgeland, A., Holme, I., Leren, P., Trygg, K.

1979. The effect of dietary changes on high density lipoprotein cholesterol. The Oslo Study. *Am. J. Med.* 66:105–9
56. Hulley, S. B., Cohen, R., Widdowson, G. 1977. Plasma high-density lipoprotein cholesterol level. Influence of risk factor intervention. *J. Am. Med. Assoc.* 238:2269–71
57. Logan, R. L., Riemersma, R. A., Thomson, M., Oliver, M. F., Olsson, A. G., Walldius, G., Rössner, S., Kaijser, L., Callmer, E., Carlson, L. A., Lockerbie, L., Lutz, W. 1978. Risk factors for ischaemic heart-disease in normal men aged 40. Edinburgh-Stockholm study. *Lancet* 1:949–55
58. Schonfeld, G., Weidman, S. W., Witztum, J. L., Bowen, R. M. 1976. Alterations in levels and interrelations of plasma apolipoproteins induced by diet. *Metabolism* 25:261–75
59. Easty, D. L. 1970. The relationship of diet to serum cholesterol levels in young men in Antarctica. *Br. J. Nutr.* 24:307–11
60. Raymond, T. L., Connor, W. E., Lin, D. S., Warner, S., Fry, M. M., Connor, S. L. 1977. The interaction of dietary fibers and cholesterol upon the plasma lipids and lipoproteins, sterol balance, and bowel function in human subjects. *J. Clin. Invest.* 60:1429–37
61. Mistry, P., Nicoll, A., Niehaus, C., Christie, I., Janus, E., Lewis, B. 1976. Cholesterol feeding revisited. *Circulation* 54:Suppl. 2, p. 178
62. Burslem, J., Schonfeld, G., Howald, M. A., Weidman, S. W., Miller, J. P. 1978. Plasma apoprotein and lipoprotein lipid levels in vegetarians. *Metabolism* 27:711–19
63. Bremner, W. F., Brooks, P. M., Third, J. L. H. C., Lawrie, T. D. V. 1975. Bran in hypertriglyceridaemia: A failure of response. *Br. Med. J.* 3:574
64. Castelli, W. P., Doyle, J. T., Gordon, T., Hames, C. G., Hjortland, M. C., Hulley, S. B., Kagan, A., Zukel, W. J. 1977. Alcohol and blood lipids. The Cooperative Lipoprotein Phenotyping Study. *Lancet* 2:153–55
65. Nikkila, E. A., Kaste, M., Ehnholm, C., Viikari, J. 1978. Increase in serum high-density lipoprotein in phenytoin users. *Br. Med. J.* 2:99
66. Helgeland, A., Hjermann, I., Leren, P., Enger, S., Holme, I. 1978. High-density lipoprotein cholesterol and anti hypertensive drugs: The Oslo Study. *Br. Med. J.* 2:403
67. Wilson, D. E., Lees, R. S. 1972. Metabolic relationships among the plasma lipoproteins. Reciprocal changes in the very-low and low-density lipoproteins in man. *J. Clin. Invest.* 51:1051–57
68. Witztum, J., Dillingham, M., Giese, W., Schonfeld, G., Weidman, S. 1978. Triglyceride lowering fails to correct low high density lipoprotein-cholesterol in primary hypertriglyceridemia (Type IV). *Circulation* 57, 58:Suppl. 2, p. 38
69. Taskinen, M.-R., Nikkila, E. A. 1979. Effects of caloric restriction on lipid metabolism in man. Changes of tissue lipoprotein lipase activities and of serum lipoproteins. *Atherosclerosis* 32:289–99
70. Garrison, R. J., Kannel, W. B., Fienlieb, M., Castelli, W. P., McNamara, P. M., Padgett, S. J. 1978. Cigarette smoking and HDL cholesterol. The Framingham Offspring Study. *Atherosclerosis* 30:17–25
71. Miller, G. J., Koplan, J. P., Morgan, P., Ashcroft, M. T., Moinuddin, M., Beckles, G. L. A. 1979. High-density lipoprotein cholesterol concentration and other serum lipids in an isolated island community free of coronary-heart disease. *Int. J. Epidemiol.* In press

Ann. Rev. Med. 1980. 31:109–26
Copyright © 1980 by Annual Reviews Inc. All rights reserved

CLINICAL PROBLEMS OF VENTILATORY CONTROL

❖7347

Gernot B. Irsigler, M.D.

Department of Medicine, H. F. Verwoerd Hospital and University of Pretoria,
Pretoria, Republic of South Africa

John W. Severinghaus, M.D.

Department of Anesthesia, University of California School of Medicine,
San Francisco, California 94143

INTRODUCTION

The concept of respiratory failure with normal lungs emerged in the 1950s from reports of patients presenting with polycythemia, hypoxemia, and hypercapnia in the absence of intrinsic cardiopulmonary disease. The pseudonyms Pickwickian Syndrome (1) and Ondine's Curse (2) helped identify, although perhaps not clarify, these patients with disordered ventilatory control to physiologists and clinicians. Recent years have seen tremendous advances in our understanding of the physiologic mechanisms involved. This subject, formerly considered the domain of a dedicated few on the "neuropulmonary fringe" (3) has decidedly entered the field of clinical medicine. Here we review normal mechanisms of ventilatory control and then discuss the integrated respiratory responses to stimuli such as hypoxia, hyperoxia, and denervation, acclimatization to high altitude, and acidosis and alkalosis. We conclude with an overview of some pathological states.

NORMAL CONTROL MECHANISMS

Breathing serves both a behavioral role (talking, laughing, etc) and a metabolic function. Voluntary breathing (behavioral) is controlled by the cortex, and automatic breathing (metabolic) by structures within the brain stem

0066-4219/80/0401-0109$01.00

with its afferent connections from chemo- and neuroreceptors. The output of these two subsystems is integrated with local reflexes from muscles and tendons to produce the definitive motor response. A schematic overview is given in Figure 1.

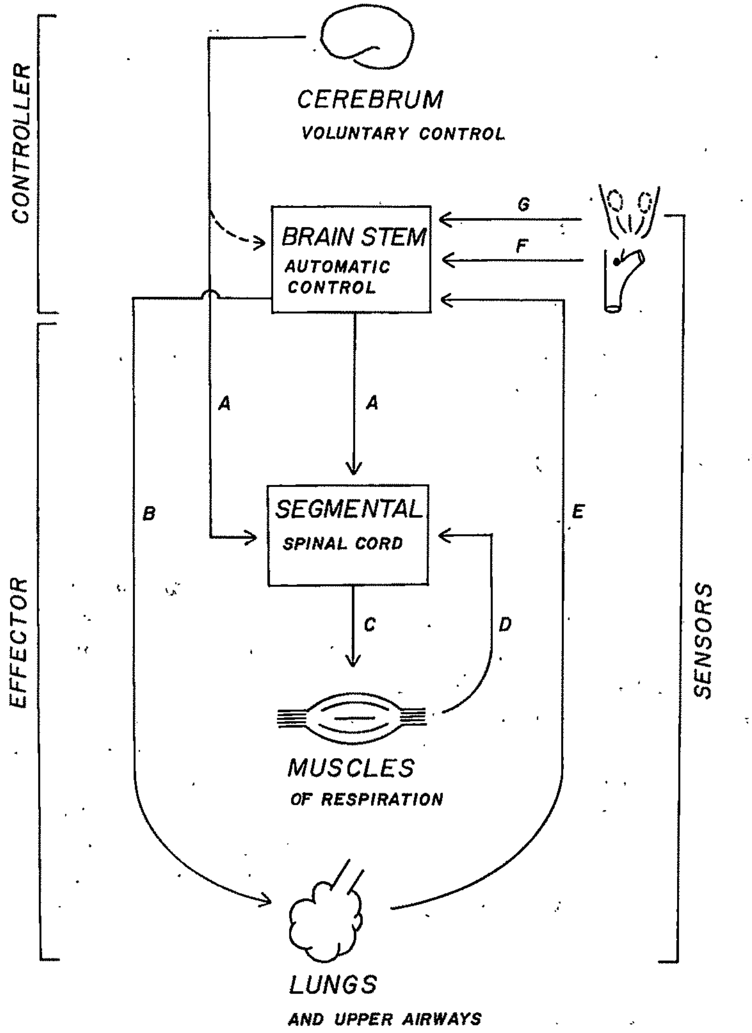


Figure 1 Schematic outline of the respiratory control system. (*A*) Spinal cord efferent pathways (voluntary and automatic). (*B*) Vagal respiratory motor efferents. (*C*) Spinal respiratory motor efferents. (*D*) Proprioceptors. (*E*) Vagal afferents (lung and upper airways). (*F*) Peripheral arterial chemoreceptors. (*G*) Central medullary chemoreceptors.

Medullary Respiratory Center and Descending Spinal Tracts

The locus for spontaneous breathing is in the brain stem medulla. The earlier concept of separate inspiratory and expiratory half centers is probably no longer tenable. Rather, two dense bilateral aggregations of respiratory neurons are involved both in initiating inspiration and expiration, and in incorporating visceral afferent impulses into the motor response (4).

The first aggregation, the dorsal respiratory group, contains primarily inspiratory neurons. These provide the motor input to the phrenic nerves, drive the ventral respiratory neurons, and probably also integrate vagal afferent impulses (from the lungs and airways) into the respiratory motor response.

A second neuron pool, the ventral respiratory group, contains both inspiratory and expiratory cells, which drive spinal motor neurons of intercostal and abdominal skeletal muscle. Respiratory rhythmicity arises in the medulla, but the nature and mechanism of the "pacemaker's" action is still under investigation (4).

Spinal Pathways and Integration

Axons from the dorsal and ventral respiratory neurons in the medulla cross the midline and descend in the ventrolateral columns of the cord. Expiratory and inspiratory tracts are spatially separated (5), as are tracts transmitting specialized reflexes (cough, hiccup) (5), and voluntary control (6). Only at the level of the segmental spinal cord are the various descending impulses integrated to produce a reflex response (4).

Peripheral Chemoreceptors

The arterial chemoreceptors, located in the carotid and aortic bodies, are responsible for the immediate increase in breathing produced by lack of oxygen. They also exert reflex systemic effects including arterial hypertension and bradycardia (7). Apart from hypoxia they respond to a fall in arterial pH. Chemoreceptor activity (and ventilation) increases exponentially as Po_2 falls below 100 mm Hg (8). Below 30 mm Hg activity gradually decreases with time (9). Hypoxic chemoreceptor stimulation is potentiated, either peripherally or centrally, by acidosis or hypercapnia. This potentiation is reflected in the amplified ventilatory response to combined hypoxia and hypercapnia. The structure and functional organization of the carotid body was recently reviewed (4) and is not discussed here.

Medullary Chemoreceptors

The medullary carbon dioxide chemoreceptors are nerve endings lying within 400 μm of the surface of the ventral medulla (10). Two distinct chemosensitive regions have been described (11, 12). Their efferent fibers form a common neuronal bundle that reaches the medullary respiratory neurons deeper down (12). The chemoreceptor stimulus is provided by H^+ ions in brain extracellular fluid (ECF). This fluid is separated from blood by the blood brain barrier, which permits free diffusion of CO_2 but not of HCO_3^- and H^+ ions. Thus CO_2 has rapid access to CSF and brain ECF, affecting its pH, whereas equilibration for HCO_3^- ions between blood and cerebrospinal fluid (CSF) is delayed.

The pH of brain ECF (and hence chemoreceptor activity) is affected by only two factors: P_{CO_2} and brain ECF HCO_3^- concentration. P_{CO_2} is governed by arterial P_{CO_2} and also by local metabolism and blood flow. Present evidence suggests a high blood-flow-to-metabolism ratio resulting in a tissue P_{CO_2} only 2–3 mm Hg above arterial P_{CO_2} (13). Brain ECF HCO_3^- concentration may be the same as that of bulk CSF since the two communicate freely.

Changes in P_{CO_2} occur rapidly following changes in arterial P_{CO_2} (time constant 60 sec). This mechanism is the short-term regulator of chemoreceptor activity. Changes in HCO_3^- concentration occur slowly (12–24 hr) and follow changes in arterial HCO_3^- concentration (metabolic acidosis/alkalosis) or P_{CO_2} (respiratory acidosis/alkalosis).

Chemoreceptor Interaction

When breathing ambient air, the balance of CO_2- and O_2-related drives heavily favors CO_2. When arterial P_{O_2} falls, at constant P_{CO_2}, the tilt is reversed: at $P_{O_2} = 40$ mm Hg (75% saturation of hemoglobin) ventilation is multiplied some three- to five-fold because of hypoxic arterial chemoreceptor stimulation. This strong response to hypoxia results in a curious phenomenon: hyperventilation, fall in arterial P_{CO_2}, and CSF alkalosis, which then eliminates most of the excess ventilatory drive (by negative feedback). In the end, ventilation is only marginally increased in the face of continued hypoxic peripheral chemoreceptor stimulation. This peripheral and central chemoreceptor interaction explains much of the CSF alkalosis at high altitude (see below) and is also consistent with the changes in CSF pH observed during experimentally induced metabolic acidosis.

Proprioceptive and Vagal Afferents

Sensory afferents from muscle and tendon receptors enter the segmental cord, augmenting motor neurons of synergistic, and inhibiting those of antagonistic, muscle groups. Vagal afferents from the lung and airways are integrated by dorsal respiratory neurons in the medulla. Activation of irritant receptors in the airways and J receptors in the lung initiate tachypnea and also reflex bronchoconstriction (14). The Hering Breuer stretch reflex, slowing respiration by inhibiting inspiration, is probably more important in animals than in man (15).

Voluntary Control

Ventilation is regulated with exquisite precision to maintain Po_2 and Pco_2 homeostasis in arterial blood, in the face of widely varying metabolic rates. This regulation is due to the automatic control system outlined above. However, the "ventilation pump" is also required for such voluntary acts as speaking, blowing, singing, etc. These acts, initiated by the motor and pre-motor cortex, temporarily suspend the otherwise continuous involuntary act of breathing. Voluntary commands project via anatomically distinct pathways to the medullary reticular formation (corticobulbar tracts) and to segmental motor effector units in the cord (corticospinal and rubrospinal tracts) (16).

Voluntary impulses may contribute to resting ventilation as "wakeful drive," and probably account for much of the individual variation in carbon dioxide and hypoxic ventilatory responses (17, 18). An interesting evolutionary adaptation has occurred in dolphins and other aquatic mammals. These animals are neurobreathers, not chemosensitive breathers: when cortical control is removed (e.g. by anesthesia), breathing stops. Their chemoreceptor responses appear to be blunted, or totally absent (19), permitting wider swings in CO_2 and O_2 tensions during "sounding." Lilly's book (20) shows a dolphin sleeping with one eye open: the primordial victim of "Ondine's Curse" (2).

Integrated Regulation of Ventilation

An important function of ventilation is to remove metabolically produced carbon dioxide. Thus the ventilation pump responds to a rise in arterial Pco_2, and also to a fall in arterial Po_2 or fall in pH, by increasing the minute volume of ventilation. These relationships are illustrated in Figure 2. Curve A, the metabolic hyperbola, expresses the relation of ventilation to arterial Pco_2 (no CO_2 added) when the effect of peripheral or central chemoreceptors is completely removed, e.g mechanical ventilation of a deeply anesthetized patient. Curve A was calculated from the CO_2 mass balance equation relating metabolic production to removal by ventilation: $\dot{V}_A \times P_aco_2 = \dot{V}CO_2 \times K$. For $\dot{V}CO_2 = 180$ ml/min and $V_D/V_T = 0.25$, this equation becomes $\dot{V}_A = 240/P_aco_2$ Curve A shows that as ventilation rises, arterial Pco_2 falls and vice versa.

Line B is the hyperoxic CO_2 response line. It was obtained by adding CO_2 to inspired air (in excess oxygen to eliminate hypoxic chemoreceptor drive) for approximately 10 minutes at each level of inspired CO_2. Ventilation increases as Pco_2 rises, and the slope of this line defines medullary chemoreceptor sensitivity to CO_2. In this example, it is 3.2 liters/min/mm Hg. Line C is the hypoxic CO_2 response line. It was obtained by rapidly lowering arterial Po_2 to 40 mm Hg (by breathing a hypoxic gas mixture) for approximately 5 minutes, at each level of inspired CO_2. Ventilation

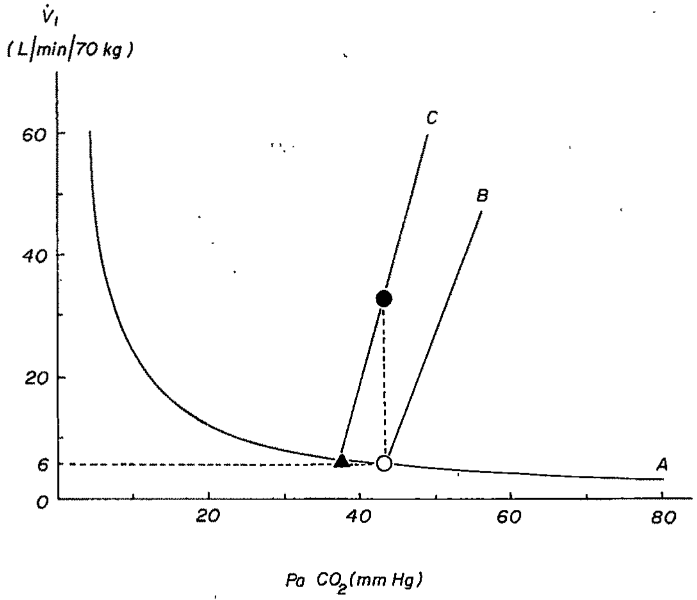


Figure 2 The relationship of ventilation to arterial carbon dioxide tension. Curve A shows the reciprocal relation between arterial carbon dioxide tension (P_aco_2) and ventilation when inspired gas contains no carbon dioxide—the metabolic hyperbola. It was constructed assuming a fixed CO_2 production (180 ml/min) and dead space tidal volume ratio 0.25. Curve B shows the steady-state ventilatory response to increased P_aco_2 with added oxygen ($P_ao_2 > 200$ mm Hg). Curve C is the same curve obtained during hypoxia ($Po_2 = 40$ mm Hg). Hypoxia steepens the ventilatory response to CO_2 with leftward displacement of the $\dot{V}_I/Pco_2$ response line. The vertical distance between Curves B and C, at any value of Pco_2, represents the ventilatory drive (in liters/min) due to hypoxia: $\dot{V}$ = inspired ventilation BTPS (body temperature and pressure, saturated) normalized to 70-kg body weight; P_aco_2 = arterial carbon dioxide tension; Open circle = Resting hyperoxic "set point" ($Pco_2 = 43.4$ mm Hg); Dashed line from open to closed circle = Hypoxic parameter $\Delta\dot{V}_{40}$ (26.4 liters/min); Closed triangle = Hypoxic "set point" ($Pco_2 = 37.2$ mm Hg). Curves B and C are mean values of five normal subjects (21).

increases as Pco_2 rises, but the slope of the line is steeper than B, and it is displaced to the left.

The vertical distance between lines B and C, at any level of Pco_2, defines the increase in ventilation due to hypoxia. In this example at $Pco_2 = 43.4$ (the resting hyperoxic "set point"), this difference is 26.4 liters/min. This is the hypoxic parameter $\Delta\dot{V}_{40}$. Lowering of arterial pH has an effect similar to that of hypoxia: metabolic acidosis of pH_a 7.30 approximately equals hypoxic stimulation of Po_2 65 mm Hg (21).

As an example of how this regulation works, let us assume that alveolar ventilation falls so that arterial Pco_2 rises by 3 mm Hg (Figure 3) (point

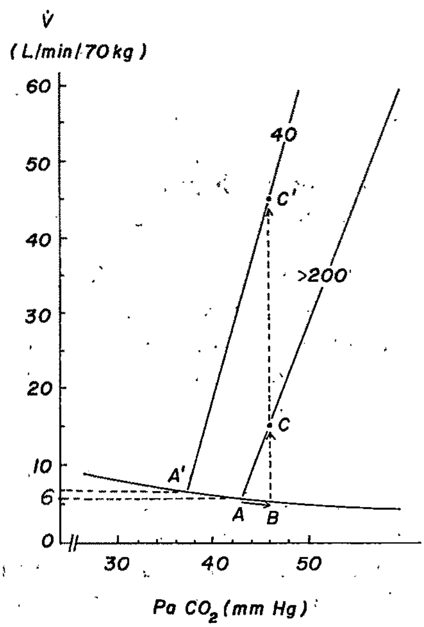


Figure 3 Mechanisms underlying ventilatory adjustment when arterial P_{CO_2} is transiently raised. This is an enlarged section of Figure 2 (see text).

A → B). If medullary chemoreceptors are now "switched on," ventilation immediately rises (to point C), thereby eliminating the excess CO_2, and ventilation returns toward set point A. If, at point B, the subject had also been made hypoxic, ($P_{O_2} = 40$ mm Hg), the combined effect of hypoxia and CO_2 would have been a much stronger stimulus to ventilation (point C'). Ventilation returns rapidly to set point A or, with continued hypoxia, to point A'. The difference in P_{CO_2} between A and A' reflects a similar difference in P_{CO_2} of brain ECF. The resultant CSF alkalosis counteracts most of the excess hypoxic drive, hence ventilation increases by only about 1.2 liters/min (at A') instead of 26.4 liters/min (normal medullary chemoreceptor pH) at A.

HYPOXIA, HYPEROXIA, AND DENERVATION

The primary oxygen sensors in man are the carotid bodies. Their influence may become paramount to ventilatory compensation when pulmonary gas exchange is impaired by disease. Chronic hypoxia occurring in patients with Fallot's tetralogy or in high altitude natives, leads to very marked blunting of the hypoxic response (22). Whether this loss of hypoxic drive is reversible is unresolved. Sørensen & Severinghaus (23) found hypoxic responsiveness markedly depressed in five patients with tetralogy of Fallot more than one year after surgical correction of the defect. In contrast Blesa et al (24) reported return of hypoxic sensitivity following surgery in some such patients.

In the absence of carotid bodies, hypoxia of 45 mm Hg reduces ventilation by a direct depressant effect on the respiratory centers (25).

Hyperoxia and Peripheral Chemoreceptor Denervation

In man, inhalation of one or two breaths of pure oxygen causes a 10% decrease in ventilation for the next several breaths (26), presumably a result of decreased carotid chemoreceptor drive. Carotid body function may be permanently lost following bilateral carotid endarterectomy for cerebrovascular ischemia (25).

In seven such patients ventilatory and blood pressure responses to hypoxia were absent following surgery. They also developed a 6-mm Hg rise in arterial P_{CO_2}, reflecting the loss of resting carotid chemoreceptor drive. The deliberate excision of carotid bodies to ameliorate asthma has now been largely discredited (27), although it is still practiced by some (28). Loss of hypoxic responsiveness is potentially serious: it converts the effect of hypoxia from stimulation to depression of ventilation, and may lead to syncope in some patients (29).

ACCLIMATIZATION PHENOMENA

Man ascending to high altitude increases his breathing immediately. Over the course of the next few hours and days, this hyperventilation increases still further and becomes established at a new steady-state value. This secondary increase in ventilation is referred to as acclimatization. The immediate increase in breathing, produced by hypoxia, is initiated and sustained by the chemoreceptors. It is abolished by adding oxygen to the inspired air ("acute normoxia") or, in animals, by prior carotid body denervation (30). In contrast, the hyperventilation in acclimatized subjects is only incompletely abolished by acute normoxia: some residual hyperventilation remains, and only with time does breathing return to its sea-level control value. The principal determinant of acclimatization is a fall in CSF HCO_3^- concentration. The mechanisms responsible for this HCO_3^- shift are still unresolved; in particular whether active ion transport (31, 32) or whether passive redistribution between CSF and blood is involved (33). Although Severinghaus et al (31) found little CSF alkalosis (despite arterial and CSF hypocapnia), more recent studies have shown that lumbar CSF pH remains substantially alkaline during acclimatization from 8 hours to 4 weeks after ascent to altitude (34, 35). Recently it was shown that this CSF alkalosis is eliminated by acute normoxia (13). Thus, at high altitude, the CSF alkalosis serves to offset the excess hypoxic peripheral chemoreceptor drive, which would increase ventilation by 6 to 15 liters/min, rather than the observed 2 to 3 liters/min (at 3810 m, $P_aO_2 = 55$ mm Hg; see below).

After three days at high altitude, arterial P_{CO_2} falls to 29 mm Hg, reflecting 2–3 liters/min hyperventilation above sea-level values. If arterial P_{O_2} is now acutely restored (by oxygen breathing), P_{CO_2} rises to 32 mm Hg, instead of the sea-level control value of 40.5 mm Hg (Figure 4). Also, the CO_2 response lines are shifted 10 mm Hg to the left (13).

What is the mechanism of this residual hyperventilation and what accounts for the shift in the CO_2 response lines? After three days at high altitude CSF HCO_3^- falls about 5 meq/liter accompanied by a nearly proportionate fall in CSF P_{CO_2} such that the central chemoreceptors are reset to a new P_{CO_2}. Hence the CO_2 response curve is shifted to the left. When

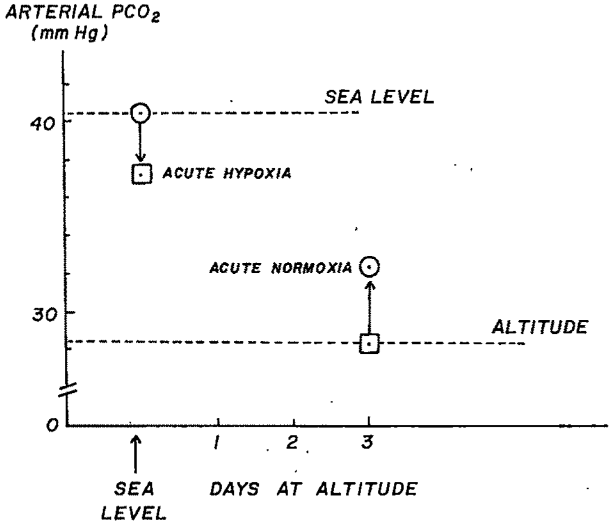


Figure 4 Effect of normoxia and hypoxia on arterial P_{CO_2} (i.e resting ventilation) (*a*) at sea level and (*b*) after acclimatization to altitude (3 days at 3810 m, ambient $P_{O_2} = 55$ mm Hg). At sea level, reduction of inspired oxygen tension to an end tidal P_{O_2} of 55 mm Hg for 10 minutes causes P_aCO_2 to fall 3.4 mm Hg, from its resting value of 40.5 mm Hg to 37.1 mm Hg. After three days at altitude, resting P_aCO_2 is 28.7 mm Hg. Acute normoxia (inspired $P_{O_2} = 100$ mm Hg) raises P_{CO_2} to 32.4 mm Hg, but not to 40.5 mm Hg, the resting sea-level value. This lower resting P_aCO_2 is due to the downward adjustment of CSF HCO_3^- concentration that has occurred at altitude. Data are from Crawford & Severinghaus (13). Open circles represent normoxia (end tidal P_{O_2} 100 mm Hg); open squares represent hypoxia (end tidal P_{O_2} 55 mm Hg).

hypoxia is now suddenly abolished, ventilation falls, P_{CO_2} rises, and the pH of CSF (and medullary chemoreceptor ECF) becomes acid, thus driving ventilation. The acid shift required to increase ventilation by 2 liters/min (i.e $P_{CO_2} = 32$ mm Hg) would be 0.005 pH units, which cannot reliably be measured by present techniques.

From the foregoing it follows that normal resting ventilation (and normal arterial P_{CO_2}) cannot be restored until CSF HCO_3^- concentration has returned to sea-level values.

These acclimatization phenomena have a clinical counterpart in patients when ventilation is driven not by hypoxic carotid chemoreceptor stimulation but by mechanical ventilation in the treatment of respiratory failure. Prolonged inadvertent hyperventilation will lower arterial and CSF P_{CO_2}, lower CSF HCO_3^- concentration, and reset medullary chemoreceptors to a lower resting arterial P_{CO_2}. Normal ventilation and normal arterial P_{CO_2} will likewise not be restored until CSF HCO_3^- concentration has once more returned to resting levels.

ACIDOSIS AND ALKALOSIS

Arterial acidosis stimulates ventilation primarily via the peripheral chemoreceptors, whether acidosis is acute or chronic. Biscoe et al (8) showed that carotid body afferent nerve traffic in the cat increased two- to three-fold when arterial pH fell from 7.45 to 7.25 at constant P_{CO_2} and P_{O_2}. In the dog, carotid body denervation delays or abolishes the ventilatory response to acidosis (36). In man, the hyperventilation of exercise-induced metabolic acidosis does not occur when carotid bodies are absent (37).

The immediate ventilatory stimulus seen with metabolic acidosis is weak because, as P_{CO_2} falls, CSF pH becomes alkaline and inhibits ventilatory drive arising from medullary chemoreceptors (21, 38, 39). As the acidosis becomes chronic (hours to days), bicarbonate concentration falls, permitting P_{CO_2} to fall proportionately. However, as in altitude acclimatization, this downward adjustment of CSF HCO_3^- concentration is incomplete and CSF remains significantly alkaline up to 48 hours after experimentally induced metabolic acidosis. During metabolic alkalosis ventilation falls and arterial P_{CO_2} rises. A "paradoxical" acid shift in CSF pH has been reported in the dog (40) and also acutely in man (38, 39). However, after 48 hours (21) and 5 days of alkalosis (38, 41) CSF pH is back to normal. This indicates that medullary chemoreceptor drive is probably also back to normal, and hence the decrease in ventilation must be due to a decrease in peripheral chemoreceptor drive.

In summary it appears that both the hyperventilation of acidosis and the hypoventilation of alkalosis are due to alterations in peripheral chemoreceptor drive. The concomitant changes in CSF pH are small, and the net effect of the CSF HCO_3^- adjustments, in both metabolic states, ensures a remarkably constant CSF pH in the face of systemic acidosis and alkalosis of wide-ranging severity.

VENTILATION DURING SLEEP

The past decade has seen a remarkable expansion in our knowledge of ventilation and respiratory control during sleep. The following is a brief overview of this fascinating subject, which has been covered more extensively in several excellent recent reviews (42, 43).

Control of Breathing During Sleep

Sleep is not a homogeneous state but rather is a cyclical pattern of sequential stages with a periodicity in human adults of about 90 minutes. Two distinct sleep states have been defined: non-rapid-eye-movement (non-REM) sleep (quiet or synchronized sleep) and rapid-eye-movement (REM) sleep (active or desynchronized sleep). During non-REM sleep, breathing is under the dominant influence of the automatic control system: responses to CO_2, hypoxia, and lung inflation are intact and are critically important for maintaining ventilation. During REM sleep, when cerebral activity is enhanced, ventilation is generally unresponsive to metabolic and reflex stimuli other than hypoxia (43).

Changes in Ventilation During Sleep

The transition from drowsiness to non-REM sleep is marked by irregular breathing with apneic periods (up to 15 sec) associated with fluctuations in arterial P_{CO_2} and P_{O_2}. Once non-REM sleep is established, the homeostatic set point is shifted to a higher threshold for CO_2, and 1–2 liters/min lower minute volume of ventilation. This results in a rise of 4–8 mm Hg in arterial P_{CO_2} and a fall of 3–10 mm Hg in P_{O_2} (43). REM sleep is characterized by an irregular pattern of breathing that may include periods of apnea of up to 20-sec duration. Alveolar ventilation alternates between hyperventilation (with hypocapnia) and hypoventilation (with hypercapnia and hypoxemia). These fluctuations in ventilation, although of little consequence in normal subjects, may have profound effects in patients with preexisting impaired gas exchange due to cardiac or pulmonary disease.

SLEEP APNEA SYNDROMES

In the past decade a number of clinical disorders have been recognized that share a common pathogenetic mechanism: recurrent periods of apnea during sleep. From a mechanistic viewpoint these disorders can be classified as central, obstructive, and complex or "mixed" (42). In central sleep apnea, airflow ceases because respiratory drive ceases—it is a true disorder of ventilatory control. In obstructive apnea, airflow ceases because upper airways close—respiratory drive is present and often increases as apnea becomes prolonged. Complex or mixed apnea typically begins with a period of central apnea followed by upper airway occlusion in the presence of gradually returning respiratory drive.

The pathological defect underlying central sleep apnea is unknown in the majority of patients. However, anatomical or functional defects of the automatic control system have been described in some. These include damage to the brain stem respiratory neurons, damage to ascending and descending spinal tracts of automatic control, or defects in carotid or medullary chemoreceptor responses to stimulation [see Phillipson (43)]. Occlusion of the airways in obstructive apnea occurs at the level of the oropharynx. Two different mechanisms have been proposed (*a*) passive inspiratory obstruction due to loss of genioglossus muscle tone (the pharyngeal abductor) combined with negative intra-airway pressure (44), and (*b*) active constriction of the airway due to muscle contraction (45). Several family members may be affected, which suggests a genetic basis in some cases (46). A common feature in many patients is a short thick neck sometimes combined with mild mandibular hypoplasia (47). Obvious anatomical causes such as enlargement of the tonsils or adenoids (47) or vocal cord paralysis (M. Plit, personal communication) may play a role in children and some adults.

Clinical Features

In adults obstructive sleep apnea most often occurs in middle-aged men (40–60 years). Typically there is a history of loud and troublesome snoring for many years. Although many patients are obese, 30% have normal or only slightly increased body weights (Table 1). In the typical patient, an apneic period develops during sleep; it lasts from 15 sec to 2 min and ends

Table 1 Symptoms and signs in 50 patients with obstructive sleep apnea[a]

Loud snoring	100%
Daytime sleepiness	78%
Intellectual deterioration	78%
Personality changes	48%
Impotence	42%
Morning headaches	36%
Hypertension	52%
Obesity	70%

[a] Adapted from Guilleminault et al (42).

with a transient lightening of the sleep stage or brief outright awakening, whereupon breathing resumes. The subject then quickly falls asleep again and the sequence of events is repeated, up to 500 times per night in severe cases. These episodes result in hypercapnia and hypoxemia with concomitant systemic and pulmonary arterial hypertension and cardiac arrhythmias. With progression, daytime manifestations appear: profound daytime sleepiness, deterioration of intellectual function, personality changes, and sexual impotence (Table 1). In time, systemic hypertension becomes sustained and a rise in hematocrit, cor pulmonale, and cardiac failure develop in response to chronic hypoxia.

In adolescents a decline in school performance may be the only manifestation of nocturnal sleep apnea. Patients with predominantly central sleep apnea are a more heterogeneous group. They complain of nocturnal insomnia, they are usually of normal weight, and they commonly are depressed (42).

Diagnosis

The diagnosis of sleep apnea syndromes may not be obvious. Routine pulmonary function tests, electrocardiogram, chest X rays and arterial blood gas measurements while awake may be completely normal. Ventilatory responses to hypoxia and to CO_2 may be subnormal, but normal responses do not exclude the diagnosis. When the diagnosis is strongly suspected on clinical grounds, a proper sleep study must be performed. The minimum aim of such a study should be to document recurrent periods of absent airflow at the nose (or mouth), and to determine whether such apneas are accompanied by continuation or absence of respiratory effort. Techniques are fully described elsewhere (42). The electrocardiogram should be monitored throughout—severe arrhythmias during sleep may be the most immediate indication for therapy.

Treatment

The treatment of sleep apnea syndromes depends on the severity of symptoms and the predominant type of apnea (obstructive or central). If an obvious anatomic airway obstruction is discovered, it should be surgically corrected. In most cases, however, no obvious anatomical cause for the sleep pathology is found. In obese patients, weight loss often reduces the incidence and severity of nocturnal apneic episodes. The most effective treatment is tracheostomy. Considerable experience with the long-term management of such patients has now been gained at Stanford University Medical Center (47).

In central apnea, a variety of medications have been used to increase respiratory drive including theophylline, acetazolamide, clomipramine, and medroxyprogesterone, with varying degrees of success. In severe cases electrical stimulation of the phrenic nerves has been employed—such pacing, however, may unmask or induce upper airway obstruction, which must also be relieved for symptomatic control (48).

Sudden Infant Death Syndrome

The sudden infant death syndrome (SIDS) accounts for an estimated 10,000 deaths in the United States each year (49). The typical patient is a generally healthy infant, 2–5 months old, who dies silently during sleep. In the past, death had been ascribed to unnatural causes such as mechanical asphyxia, but few continue to hold this view today. In recent years there is increasing evidence to suggest that SIDS may be a disease of disordered respiratory control.

Post mortem studies show evidence of chronic alveolar hypoventilation (50), yet the airways and bellows mechanism of the chest are normal. Carotid body tissue is hypoplastic in some (51) and there is abnormal brain stem astroglial proliferation in others (52). However, these changes may be the result of chronic hypoxemia, not its cause (52). A recent report has linked an infant's death, presumably due to SIDS, with adult sleep apnea occurring in other members of the same family (46).

In infants REM sleep predominates during the first two months of life, and non-REM sleep increases during later months (43). Apnea during REM sleep can be triggered by upper airway reflexes (53) and several authors have noted an increased incidence of upper respiratory infection in victims of SIDS (54). Ventilation in non-REM sleep is critically dependent on intact automatic respiratory control (55). There is evidence in some victims of impaired ventilatory response to CO_2 (56), and prolonged apnea has been clearly associated with non-REM sleep in one recent report (57).

In summary, there is much circumstantial evidence to suggest that death in SIDS is due to defective respiratory control. The physiological mechanisms leading to apnea and death may differ depending on the infant's age. The site and nature of the defect also remain to be determined.

Pickwickian Syndrome and Ondine's Curse

In 1956 Burwell and co-workers (1) suggested the term Pickwickian Syndrome to describe the symptom complex of obesity, daytime somnolence, alveolar hypoventilation, and cardiac failure, in patients who have no underlying pulmonary or cardiac disease. An essential feature of the obesity hypoventilation syndrome is obstructive sleep apnea, which was not recognized in earlier case reports (58). It is now known that only a minority of obese patients hypoventilate. Furthermore many hypoventilating patients with normal lungs are not obese (42).

The term Ondine's Curse was used by Severinghaus & Mitchell (2) to describe a patient whose ninth (and probably tenth) nerve roots had been damaged by surgery. This patient was left with impaired automatic ventilatory control: when unattended for only a few minutes, she would literally forget to breathe. Although the term has subsequently been widely used (59) to describe loss of automatic ventilatory drive, the more appropriate general designation of "sleep apnea" should probably replace both of these terms.

The Impact of Chronic Lung Disease

Lung disease imposes mechanical loads that diminish movements of the lung pump in response to normal respiratory center drive. These loads are of two kinds: (*a*) elastic as in congestion, atelectasis, fibrosis, or chest wall deformities, and (*b*) resistive as in all forms of obstructive airways disease. Lung disease also impairs the efficiency of gas exchange such that the normal minute volume of ventilation no longer suffices to keep P_{CO_2} normal.

Diseases imposing elastic loads induce tachypnea and hyperventilation, which is initiated both by chemical (carotid chemoreceptor) and neural afferent impulses (vagal afferents from the lung) (60).

Acutely imposed resistive loads may also cause hyperventilation. Many patients with asthma have a low arterial P_{CO_2}, even when breathing oxygen. Presumably, excess vagal afferent stimuli are likewise involved. Chronic obstructive airways disease leads to hypoventilation and CO_2 retention in some patients, but not in others. The reason for this is unresolved. Clearly mechanical limitation is not the only factor, since indices of airflow obstruction correlate poorly with CO_2 retention (61). Lung disease in these patients may unmask a preexisting diminished CO_2 responsiveness, by imposing a mechanical load (62). Others have proposed diminished carotid chemoreceptor drive due to chronic hypoxia (63). Neither of these hypotheses is supported by a recent observation that resting inspiratory drive is increased in patients with obstructive airways disease, whether hypercapnic or not (64).

When arterial and brain ECF P_{CO_2} rise acutely, the resulting acidosis causes a powerful increase in respiratory drive both from carotid and medullary chemoreceptors. If hypercapnia persists (subacute to chronic respiratory acidosis) arterial pH, and thus carotid chemoreceptor drive, is restored toward normal by renal HCO_3^- retention (65). Similarly the rise in HCO_3^- concentration in bulk CSF (and presumably brain ECF surrounding medullary chemoreceptors) shifts medullary chemoreceptor pH slightly toward normal, reducing their drive (65, 66). This leaves the carotid body response to hypoxia as the important determinant of ventilatory drive, and accounts for the well-known hazard of uncontrolled oxygen administration in these patients.

Chronically hypoxic patients are particularly vulnerable during sleep. As discussed earlier, the deeper stages of non-REM sleep are accompanied by a fall in arterial P_{O_2} (and rise in P_{CO_2}), which in these patients may reach critical levels. Also, the normally brief apneas of REM sleep may result in profound arterial desaturation (and increased pulmonary artery pressure) because preëxisting hypoxemia places these patients on the steeper part of the hemoglobin dissociation curve. Thus patients whose degree of pulmonary hypertension or cardiac failure are out of proportion to their arterial blood gas values while awake may have significant nocturnal hypoxemia. The judicious use of low flow oxygen at night may be of particular benefit in such cases.

Literature Cited

1. Burwell, C. S., Robin, E. D., Whaley, R. D. et al. 1956. Extreme obesity associated with alveolar hypoventilation: a Pickwickian Syndrome. *Am. J. Med.* 21:811–18
2. Severinghaus, J. W., Mitchell, R. A. 1962. Ondine's Curse—failure of respiratory center automaticity while awake. *Clin. Res.* 10:122
3. Butler, J. 1974. Clinical problems of disordered respiratory control. *Am. Rev. Resp. Dis.* 110:695–98
4. Berger, A. J., Mitchell, R. A., Severinghaus, J. E. 1977. Regulation of respiration. *N. Engl. J. Med.* 297:92–97
5. Newsom Davis, J., Plum, F. 1972. Separation of descending spinal pathways to respiratory motor neurones. *Exp. Neurol.* 34:78–94
6. Aminoff, M. J., Sears, T. A. 1971. Spinal integration of segmental, cortical, and breathing inputs to thoracic respiratory motor neurones. *J. Physiol. (London)* 215:557–75
7. Comroe, J. H. Jr. 1964. The peripheral chemoreceptors. In *Handbook of Physiology, Respiration,* ed. W. O. Fenn, H. Rahn, pp. 557–83. Washington, DC: Am. Physiol Soc.
8. Biscoe, T. J., Purves, M. J., Sampson, S. R. 1970. The frequency of nerve impulses in single carotid body chemoreceptor afferent fibres recorded in vivo with intact circulation. *J. Physiol. (London)* 208:121–31
9. Lahiri, S., de Laney, R. G. 1975. Stimulus interaction in responses of carotid body chemoreceptor single afferent fibres. *Respir. Physiol.* 24:249–66
10. Berndt, J., Berger, W., Mückenhoff, K. 1972. Untersuchungen zum zentralen chemosensiblen Mechanismus der Atmung. *Pfluegers Arch.* 332:127–45
11. Mitchell, R. A., Loeschke, H. H., Massion, W. H. et al. 1963. Respiratory responses mediated through superficial chemosensitive areas on the medulla. *J. Appl. Physiol.* 18:523–33
12. Schlaefke, M. E., See, W. R., Loeschke, H. H. 1970. Ventilatory response to alterations of H^+ ion concentration in small areas of the ventral medullary surface. *Respir. Physiol.* 10:198–212
13. Crawford, R. D., Severinghaus, J. W., 1978. CSF pH and ventilatory acclimatization to altitude. *J. Appl. Physiol. Respir. Environ. Exer. Physiol.* 45(2): 275–83
14. Widdicombe, J. G. 1974. Reflex control of breathing. In *Respiratory Physiology: MTP International Review of Science,* ed. J. G. Widdicombe, Ser. 1, Vol. 2, pp. 273–302, London: Butterworth
15. Phillipson, E. A. 1974. Vagal control of breathing pattern independent of lung inflation in conscious dogs. *J. Appl. Physiol.* 37:183–89
16. Mitchell, R. A., Berger, A. J. 1975. Neural regulation of respiration. *Am. Rev. Respir. Dis.* 111:206–24
17. Irsigler, G. B. 1976. Carbon dioxide response lines in young adults: the limits of the normal response. *Am. Rev. Respir. Dis.* 114:529–36
18. Sahn, S. A., Zwillich, C. W., Dick, N. et al. 1977. Variability of ventilatory responses to hypoxia and hypercapnia. *J. Appl. Physiol. Respir. Environ. Exer. Physiol.* 43(6):1019–25
19. Plum, F. 1974. Cerebral control of breathing. In *Ventilatory and Phonatory Control Systems,* ed. B. Wyke. Oxford University Press. 220 pp.
20. Lilly, J. C. 1961. *Man and Dolphin.* Garden City, NY: Doubleday

21. Irsigler, G. B., Stafford, M. J., Severinghaus, J. W. 1979. Relationship of CSF pH, O_2, and CO_2 responses in metabolic acidosis and alkalosis in humans. *J. Appl. Physiol. Respir. Environ. Exer. Physiol.* In press
22. Sørensen, S. C., Severinghaus, J. W. 1968. Irreversible respiratory insensitivity to acute hypoxia in man born at high altitude. *J. Appl. Physiol.* 25(3): 217–20
23. Sørensen, S. C., Severinghaus, J. W. 1968. Respiratory insensitivity to acute hypoxia persisting after correction of tetralogy of Fallot. *J. Appl. Physiol.* 25(3):221–23
24. Blesa, M. I., Lahiri, S., Rashkind, W. J. et al. 1977. Normalization of the blunted ventilatory response to acute hypoxia in congenital cyanotic heart disease. *N. Engl. J. Med.* 296:237–41
25. Wade, J. G., Larson, C. P., Hickey, R. F. et al. 1970. Effect of carotid endarterectomy on carotid chemoreceptor and baroceptor function in man. *N. Engl. J. Med.* 282:823–29
26. Dejours, P. 1963. Control of respiration by arterial chemoreceptors. *Ann. NY Acad. Sci.* 109:682–95
27. Curran, W. S., Graham, W. G. B. 1971. Long term effects of glomectomy. *Am. Rev. Respir. Dis.* 103:566–68
28. Winter, B. 1972. Bilateral carotid body resection for asthma and emphysema. *Int. Surg.* 57:458–66
29. Chang, K. C., Morrill, C. G., Chai, H. 1978. Impaired response to hypoxia after bilateral carotid body resection for treatment of bronchial asthma. *Chest* 73(5):667–68
30. Bouverot, P., Bureau, M. 1975. Ventilatory acclimatization and CSF acid base balance in carotid chemodenervated dogs at 3550 m. *Pfluegers Arch.* 361: 17–23
31. Severinghaus, J. W., Mitchell, R. A., Richardson, B. W. et al. 1963. Respiratory control at high altitude suggesting active transport regulation of CSF pH. *J. Appl. Physiol.* 18:1155–66
32. Bledsoe, S. W., Mines, A. H. 1975. Effect of plasma [K^+] on the DC potential and on ion distribution between CSF and blood. *J. Appl. Physiol.* 39:1012–16
33. Pavlin, E. G., Hornbein, T. F. 1975. Distribution of H^+ and HCO_3^- between CSF and blood during metabolic acidosis in dogs. *Am. J. Physiol.* 228:1134–40
34. Dempsey, J. A., Vorster, H. V., do Pico, G. A. 1974. Ventilatory acclimatization to moderate hypoxemia in man. *J. Clin. Invest.* 53:1091–1100
35. Forster, H. V., Dempsey, J. A., Chosy, L. W. 1975. Incomplete compensation of CSF [H^+] in man during acclimatization to high altitude (4300 m). *J. Appl. Physiol.* 38:1067–72
36. Bainton, C. R. 1978. Canine ventilation after acid-base infusions, exercise, and carotid body denervation. *J. Appl. Physiol. Respir. Environ. Exer. Physiol.* 44(1):28–35
37. Wasserman, K., Whipp, B. J. 1976. The carotid bodies and respiratory control in man. In *Morphology and Mechanisms of Chemoreceptors*, ed. A. S. Paintal, pp. 156–74. New Delhi: Nauchetan
38. Bradley, R. D., Semple, S. J. G. 1962. A comparison of certain acid-base characteristics of arterial blood, jugular venous blood and cerebrospinal fluid in man, and the effect on them of some acute and chronic acid-base disturbances. *J. Physiol. (London)* 160: 381–91
39. Mitchell, R. A., Singer, M. M. 1965. Respiration and cerebrospinal fluid pH in metabolic acidosis and alkalosis. *J. Appl. Physiol.* 20:905–11
40. Robin, E. D., Whaley, R. D., Crump, C. H. et al. 1958. Acid-base relations between spinal fluid and arterial blood with special reference to control of ventilation. *J. Appl. Physiol.* 13:385–92
41. Fencl, V., Vale, J. R., Broch, J. A. 1969. Respiration and cerebral blood flow in metabolic acidosis and alkalosis in humans. *J. Appl. Physiol.* 27:67–76
42. Guilleminault, C., Van den Hoed, J., Mitler, M. 1978. Clinical overview of the sleep apnea syndromes. In *Sleep Apnea Syndromes*, ed. C. Guilleminault, W. C. Dement, pp. 1–12. New York: Liss
43. Phillipson, E. A. 1978. Control of breathing during sleep. *Am. Rev. Respir. Dis.* 118:909–39
44. Remmers, J. E., de Groot, W. J., Sauerland, E. K. 1978. See Ref. 42, pp. 211–17
45. Weitzman, E. D., Pollak, C. O., Borowiecki, B. et al. 1978. See Ref. 42, pp. 235–48
46. Strohl, K. P., Saunders, N. A., Feldman, N. T. et al. 1978. Obstructive sleep apnea in family members. *N. Engl. J. Med.* 299:969–73
47. Hill, M. W., Simmons, F. B., Guilleminault, C. See Ref. 42, pp. 347–52
48. Glenn, W. W. L., Gee, J. B. L., Cole, D. R. et al. 1978. Combined central hypoventilation and upper airway obstruction. *Am. J. Med.* 64:50–60
49. Bergman, A. B., Ray, C. G., Pomeroy, M. A. et al. 1972. Studies of the sudden

infant death syndrome in King County Washington. III Epidemiology. *Pediatrics* 49:860–70
50. Naeye, R. L. 1973. Pulmonary arterial abnormalities in the sudden infant death syndrome. *N. Engl. J. Med.* 289: 1167–70
51. Naeye, R. L., Fisher, R., Ryser, M. et al. 1976. Carotid body in the sudden infant death syndrome. *Science* 191: 567–69
52. Naeye, R. L. 1976. Brain stem and adrenal abnormalities in the sudden infant death syndrome. *Am. J. Clin. Pathol.* 66:526–30
53. Sutton, D., Taylor, E. M., Lindeman, R. C. 1978. Prolonged apnea in infant monkeys resulting from stimulation of superior laryngeal nerve. *Pediatrics* 61:519–27
54. Steinschneider, A. 1972. Prolonged apnea and the sudden infant death syndrome: clinical and laboratory observations. *Pediatrics* 50:646–54
55. Phillipson, E. A., Sullivan, C. E. 1978. See Ref. 42, pp. 47–64
56. Shannon, D. C., Kelly, D. H., O'Connell, K. 1977. Abnormal regulation of ventilation in infants at risk for sudden infant death syndrome. *N. Engl. J. Med.* 297:747–50
57. Shannon, D. C., Marsland, D. W., Gould, J. B. et al. 1976. Central hypoventilation during quiet sleep in two infants. *Pediatrics* 57:342–46
58. Gastaut, H., Tassinari, C. A., Duron, B. 1966. Polygraphic study of episodic diurnal and nocturnal (hypnic and respiratory) manifestations of the Pickwick Syndrome. *Brain Res.* 2:167
59. Sugar, O. 1978. In search of Ondine's Curse. *J. Am. Med. Assoc.* 240:236–37
60. Phillipson, E. A., Murphy, E., Kozar, L. F. et al. 1975. Role of vagal stimuli in exercise ventilation in dogs with experimental pneumonitis. *J. Appl. Physiol.* 39:76–85
61. Clark, T. J. H., Freedman, S., Campbell, E. J. M. et al. 1969. The ventilatory capacity of patients with chronic airways obstruction. *Clin. Sci.* 36:307–16
62. Howell, J. B. L. 1974. Loaded breathing. In *Proc. Int. Symp. on the Effects of Mechanical Loads on Breathing,* 12–14 April 1973, McMaster Univ. Hamilton, Canada, ed. L. D. Peugelly, A. S. Rebuck, E. J. M. Campbell, p. 6. New York: Longman
63. Kepron, W., Cherniack, R. M. 1973. The ventilatory response to hypercapnia and to hypoxemia in chronic obstructive lung disease. *Am. Rev. Respir. Dis.* 108:843–50
64. Grassino, A., Sörli, J., Lorange, G. et al. 1978. Respiratory drive and timing in chronic obstructive pulmonary disease. *Chest* 73: Suppl., pp. 290–93
65. Bleich, H. L., Berkman, P. M., Schwartz, W. B. 1964. The response of cerebrospinal fluid composition to sustained hypercapnia. *J. Clin. Invest.* 43(1):11–16
66. Alroy, G. G., Flenley, D. C. 1967. The acidity of the cerebrospinal fluid in man with particular reference to chronic ventilatory failure. *Clin. Sci.* 33:335–43

Ann. Rev. Med. 1980. 31:127–37
Copyright © 1980 by Annual Reviews Inc. All rights reserved

FLATULENCE

❖7348

Michael D. Levitt, M.D. and John H. Bond, M.D.

Department of Medicine, Minneapolis Veterans Administration Medical Center,
and the University of Minnesota, Minneapolis, Minnesota 55455

"Doctor, I have too much gas," is one of the most common complaints heard by generalists and, perhaps, the most common complaint of patients consulting gastroenterologists. Among the various symptoms attributed to excessive gas are belching, bloating, a variety of abdominal pains, and excessive flatulence. Although the exact role of excessive bowel gas in the above symptom complex has not been totally defined, recent studies suggest that gas is usually a coincident occurrence that has been incorrectly impugned virtually from time immemorial (1).

The role of bowel gas in clinical symptomatology is best discussed if the reader understands the factors influencing the accumulation and elimination of gas from the gut. Therefore, this chapter first discusses the physiology of intestinal gas and then the application of this physiology to the various clinical "gas" syndromes.

PHYSIOLOGY OF INTESTINAL GAS

Origin of Bowel Gas

Five gases, N_2, O_2, CO_2, H_2, and CH_4 account for over 99% of bowel gas (2). These gases can be derived from three different sources: (*a*) swallowed air; (*b*) production within the gut; and (*c*) diffusion from the blood into the gut lumen.

SWALLOWED AIR Air can be swallowed in the form of gas trapped in food (an apple is alleged to be 20% gases by volume); in association with food ingestion (drinking a glass of water can result in the deposition of two times as much air as water in the stomach); and by the nervous habit of swallowing or aspirating air unassociated with food ingestion.

0066-4219/80/0401-0127$01.00

Although many patients are clinically labeled as air swallowers, the studies of Maddock and co-workers some 30 years ago appear to provide the only quantitative data on this phenomenon (3, 4). According to these investigators, air swallowing results from the production of a negative intraesophageal pressure in association with relaxation of the upper esophageal sphincter. Thus, air "sucking" is a more accurate description of the process. When the superior esophageal sphincter was short-circuited via a nasogastric tube in the esophagus, each inspiration against a closed glottis moved large amounts of air into the esophagus. Most of the gas was eructated; however, from 50 to 160 ml per aspirate could end up in the stomach. Clinical air swallowers appear to be individuals who consciously or unconsciously are capable of relaxing the upper esophageal sphincter during inspiration.

An unusual form of "air sucking" was practiced by the infamous Frenchman, Joseph Pujol (stage name: Le Petomane), who entertained at the Moulin Rouge by "playing" various tunes via the passage of apparently unlimited quantities of rectal gas. Monsieur Pujol had the unique ability to consciously relax his anal sphincter and simultaneously reduce the intraabdominal pressure, thus permitting aspiration of gas (or liquid) into the colon.

GAS PRODUCTION IN THE GUT Gas can be produced in the gut when acid reacts with bicarbonate to release CO_2, or it can result from direct elaboration by intestinal bacteria. Reaction of 1 meq of acid with bicarbonate produces 22.4 ml of CO_2. The acid for such a reaction can be derived from several sources. In vitro titration of gastric acid secretion indicates that the stomach normally secretes about 30 meq per hour of hydrogen ion in response to a meal. In addition, acid is released when ingested lipid, largely in the form of triglycerides, is digested to fatty acid by pancreatic lipase. Such digestion releases about 100 meq of fatty acids from the 30 g of fat ingested in a fairly hearty meal. Thus, theoretically, as much as 200 meq of acid could be released in the upper small bowel in the several-hour period following a meal and, if completely neutralized by bicarbonate in the lumen, this acid could result in the liberation of about 3000 ml of CO_2.

A second source of acid in the gut is the volatile organic acids such as acetic or propionic acid produced by bacteria during fermentation reactions. The copious quantities of CO_2 excreted after ingestion of foods containing nonabsorbable carbohydrates [such as beans (5, 6)] presumably results in part from this mechanism.

The second source of gas production in the bowel is the bacterial production of volatile metabolites. While H_2, CO_2, and CH_4 are the only such metabolites that represent an appreciable fraction of bowel gas, trace quantities of volatiles produced by bacteria, such as skatole and H_2S, are the source of the noxious odor of flatus.

There is no evidence that human cellular metabolism can produce hydrogen; apparently all of the H_2 elaborated in man results from the metabolism of gut bacteria. This concept is supported by studies in gnotobiotic rats (7) and human newborns (8) that showed no H_2 production in the germ-free state but copious H_2 production after contamination with bacteria.

Using intestinal intubation to distinguish between H_2 production in the small and large bowel, we found virtually all H_2 production in normal subjects occurred in the colon. This colonic H_2 production was very low in the fasting state but increased markedly when a fermentable substrate, lactose, was infused into the small bowel and colon (9). Thus, it appears that H_2-producing bacteria are normally limited to the colon and the delivery to the colonic bacteria of ingested fermentable substrate (as opposed to endogenous substrates) is required to support H_2 production. Therefore, the major requirement for H_2 production in the gut is the ingestion of carbohydrates that are not entirely absorbed in the small bowel. The most common form of intestinal disease resulting in such carbohydrate malabsorption is lactase deficiency, which results in lactose malabsorption. The massive quantities of gas produced after ingestion of beans appears largely attributable to the high concentrations of oligosaccharides such as stachyose and raffinose in legumes (5), which cannot be digested by enzymes in the normal gastrointestinal tract and thus pass unabsorbed through the small bowel. These carbohydrates are fermented by the colon bacteria with the direct release of H_2 and CO_2; in addition, organic acids are produced that react with bicarbonate to produce additional CO_2.

Methane, like H_2, is produced solely by bacteria (methanogenic bacteria) in the colon (10). Unlike H_2, CH_4 production in man does not appear to depend on the ingestion of nonabsorbed substances, and the substrates necessary for CH_4 production have not been clearly defined. In addition, CH_4 production differs from that of H_2 in that only about one third of adult Americans possess sufficient numbers of CH_4-liberating bacteria to produce appreciable quantities of this gas (11). This tendency to possess a CH_4-producing flora is a relatively constant individual trait and subjects studied repeatedly at intervals over a one-year period maintain the tendency to produce or not produce CH_4. The possession of a CH_4-producing flora appears to be an environmentally acquired, familial trait (11).

DIFFUSION OF GASES FROM BLOOD AND LUMEN The intestinal gases are highly soluble and diffuse rapidly between blood and lumen, the net movement of a gas being down the partial pressure gradient that exists between blood and lumen. The only three gases present in appreciable concentration in the blood (and hence, the only gases that might diffuse from the blood to the lumen) are N_2, O_2, and CO_2. Since CO_2 is produced in the gut, the P_{CO_2} in the lumen is usually higher than in the blood. Thus, CO_2 usually diffuses from lumen to blood, and such diffusion is not a source of intestinal gas.

Elimination of Gas From the Gut

Gas can be removed from the bowel lumen in three ways: convection via eructation and passage per rectum, metabolism by the intestinal bacteria, and diffusion into the blood.

ERUCTATION AND PASSAGE PER RECTUM Eructation has not been quantitated but there appear to be individual differences in the ease with which the gastric air bubble is eliminated via this mechanism. Surgical correction of hiatus hernia may lead to total inability to eructate (12). In addition, eructation may be influenced by posture. The esophagus enters the superior-posterior aspect of the stomach, thus reclining supine causes a fluid layer to form over the gastric esophageal junction and thereby prevents belching (13). The frequent development of rapid, massive gaseous distension of the intestine during surgical and X-ray procedures performed in the supine position probably results from the combination of nervous air sucking and a posture that favors movement of air into the duodenum.

Passage of gas per rectum is a much discussed social phenomenon that has received somewhat less scientific study. There appear to be wide individual variations in the composition and volume of gas passed per rectum (from 400 to 2400 ml per day) (14). The number of passages of gas per day in young, normal males averaged about 14 ± 4 (1 SD) (15).

UTILIZATION OF GAS BY BACTERIA Some colonic bacteria rapidly consume H_2 liberated by other organisms in the gut (16). When studied in vitro in the form of fecal homogenates, this consumption can proceed at an extremely rapid rate (approximately 2 ml/hr·g feces). When H_2 was infused into the colon of the rat, over 90% was consumed by colonic bacteria and about 10% was excreted (16). Thus, H_2 consumption plays as important a role as does H_2 production in determining the amount of H_2 available for excretion.

Oxygen is also rapidly consumed by colonic bacteria, and O_2 concentration in flatus is nearly always low (< 30 mm Hg). A high O_2 concentration in a flatus sample can be used as evidence of atmospheric contamination.

DIFFUSION FROM THE GUT As discussed previously, a gas will be absorbed from the gut into the blood if the partial pressure of gas in the gut exceeds that in the blood. The partial pressure of CH_4 and H_2 in the gut is always higher in the lumen than in blood. Once absorbed these gases are carried to the lung where they are efficiently cleared in a single passage and excreted in expired air. Analysis of expired air thus provides a simple means of semi-quantitatively measuring the rate of production of these gases in the lumen (17).

Composition of Intestinal Gas

The composition of gas in any given location in the gut represents the net result of the above described production and elimination processes in the intestinal tract proximal to that location. The gastric gas bubble usually is largely N_2 and O_2 because it comes from swallowed air rather than from production within the lumen. Duodenal gas has a very high P_{CO_2} ranging from 200 to 500 mm Hg (18). Thus, CO_2 formed by the reaction of acid and bicarbonate represents a major source of duodenal gas. Gas composition in the more distal small bowel has not been measured. The composition of colonic gas, as evidenced by flatus composition, is highly variable. Hydrogen and CO_2 tend to predominate in periods of high flatus production following ingestion of nonabsorbed carbohydrates, while N_2 predominates during periods of low flatus output (5). The N_2 in flatus is commonly considered to be derived solely from swallowed air. However, it is also possible for N_2 to diffuse from the blood into the lumen because the production of gases in the lumen (CO_2, CH_4, and H_2) reduces the P_{N_2} of the lumen to a value below that of blood (2).

The gas in the fasting intestinal tract of ten normal subjects was "washed out" (and collected) by means of a rapid infusion of argon into the duodenum, and then studied to determine volume and composition (2). The normal gut contained about 200 ml of gas. The composition was highly variable: N_2 predominated in seven subjects, while the gases produced in the gut (H_2, CO_2, and CH_4) represented more than 50% of the bowel gas in three subjects.

CLINICAL PROBLEMS RELATED TO INTESTINAL GAS

Patients' complaints of intestinal gas commonly take one of three forms: excessive or repetitive belching, abdominal discomfort and bloating, or passage of excessive flatus from the anus. Additional clinical problems in which intestinal gas may contribute to symptoms include acute gastric dilatation, the "gas-bloat syndrome" following total laryngectomy or surgery for gastroesophageal reflux, mechanical bowel obstruction and intestinal ileus, and explosions during electrosurgery in the colon.

Repetitive Eructation

Patients who complain of chronic and repetitive belching have been observed to precede each eructation with a swallow or aspiration of air into the esophagus (3, 4). Most of this air passes only part way down the esophagus and then is noisily expelled. Thus, repetitive eructation represents a nervous habit often exacerbated by emotional upsets. Occasionally, such belching is precipitated by other symptoms of organic disease in the abdomen or thorax. In this situation, the patient apparently believes that belching will relieve the discomfort. A vicious cycle results in which pain leads to air "swallowing" which leads to increased discomfort. The patient, unaware he is swallowing air, perceives this process as an abnormality of digestion causing increased gas production, and this belief further heightens his anxiety and perpetuates his symptoms.

Since chronic belching may occasionally be a secondary symptom of organic disease, patients with this complaint first require appropriate evaluation to rule out occult thoracic or abdominal pathology. If none is detected, treatment then consists mainly of forceful reassurance that no serious disease is responsible for their belching. The patient should be told and shown that each belch simply represents the loud expulsion of air that has just been aspirated into the esophagus. The physician might demonstrate using a mirror and palpation of the neck that each eructation is preceded by an elevation of the larynx allowing aspiration of air into the esophagus. A physician who is able to belch at will can provide a convincing personal demonstration of this aerophagia-eructation sequence. Even if patients continue to belch following this simple approach, they usually cease complaining or worrying about this gas symptom once they realize its benign, functional nature.

Abdominal Discomfort and Bloating

Abdominal discomfort and bloating, often worse after eating, are probably the complaints most frequently attributed to intestinal gas. Both patients and their physicians commonly believe such symptoms are due to excessive volumes of intestinal gas, although little scientific data exists supporting this conclusion. In fact, a recent study measured the total intestinal gas volume in patients with these complaints, and failed to demonstrate increased gas either in the fasting state (19) or following a large meal (20). The patients did, however, appear to have an abnormality of intestinal motility resulting in the retarded passage of gas through the bowel and increased reflux of duodenal gas back into the stomach. In addition, they often sensed pain in the presence of intestinal gas volumes well tolerated by normal, control subjects. Thus, symptoms produced by disordered motility and pain response to gut distension seem to be misinterpreted as a feeling of increased gas when the total intestinal gas volume is actually normal. The belief by many of these patients that certain foods increase the volume of gas may represent the tendency for these foods to enhance this abnormal gut motility rather than their tendency to increase intestinal gas production.

The initial evaluation of patients with functional abdominal pain and bloating should usually include screening laboratory tests, X-ray studies of the upper gastrointestinal tract and colon, and proctosigmoidoscopy to rule out the presence of organic disease. For example, patients with colonic carcinoma not infrequently present with vague abdominal discomfort and distension. If such studies are normal, initial treatment consists of strong reassurance that no serious or progressive pathology exists. A discussion of gut motility and the relation of emotions to this motility is also helpful. This approach alone often greatly increases the patient's tolerance of abdominal sensations even though they continue to exist. Relief of apprehension about the origin of these symptoms may also favorably influence the motor function of the gut and thereby decrease the frequency or severity of functional abdominal pain.

Since these patients have a derangement in intestinal motor function, drugs that influence motility theoretically may relieve symptoms. Though several anticholinergic agents are claimed effective, controlled studies fail to show conclusively that significant benefits result from these medications. A trial course of one of the several available anticholinergic drugs might be tried in patients who continue to have severe symptoms, but the drug should be stopped if no improvement occurs within a short time.

While these patients do not have excessive intestinal gas, it appears that volumes of gas normally well tolerated cause discomfort. Thus, it may be helpful to reduce the quantity of gas accumulating in the bowel as a result of both air swallowing and intraluminal intestinal gas production. The patient should avoid repetitive belching since a portion of the aspirated air may enter the stomach, and should avoid reclining immediately after meals. Postprandial production of CO_2 in the duodenum may be decreased by both a low fat diet and administration of a hydroxide-containing antacid 30 minutes after every meal. Since both CO_2 and H_2 production by the colonic bacteria require a fermentable substrate, the patient should decrease ingestion of foods that contain large amounts of nonabsorbable carbohydrates (e.g. beans, cabbage, apples). In addition, since the sugar in milk, lactose, is not completely absorbed even by subjects without obvious lactase deficiency (21), a lactose-free diet might be tried.

Excessive Passage of Flatus

The passage of large volumes of gas from the anus may just be an annoying source of embarrassment to the patient or it may cause him to believe something is seriously wrong with his digestive system. Excessive excretion of flatus can most frequently be related to the composition of the diet and virtually never indicates serious organic disease of the gastrointestinal tract. In general, it appears that complaints of excessive passage of flatus are more closely related to the psychosocial factors than the actual quantity of gas excreted.

Conventional blood, X-ray, and endoscopic studies are of no use in the evaluation of the patient complaining of excessive flatus. A more rational approach is to first document the presence of increased flatus by inserting a rectal tube attached to a syringe and collecting all gas excreted over several hours. Normal subjects pass less than 100 ml per hour (14). If this is not feasible, the frequency of gas passage may serve as a rough estimate of total flatus elimination. Normal subjects pass gas less than 25 times per day (15). If excessive gas excretion is found, gas chromatographic analysis of a sample of rectal gas will determine if it is the result of intraluminal gas production or simply due to air swallowing. If swallowed air is the cause, rectal gas will contain high concentrations of N_2. Since increased air swallowing is usually psychogenic, simple counseling and reassurance is the only treatment required.

If rectal gas analysis shows high concentrations of H_2 and CO_2, gases largely derived from colonic fermentation of unabsorbed dietary carbohydrate, the patient should be evaluated for a malabsorptive disorder. Carbohydrate malabsorption may be caused by a generalized malabsorptive problem or may result from an isolated abnormality such as lactase deficiency. Increased production of H_2 and CO_2 in the colon of patients without malabsorption may also be due to ingestion of fruits, vegetables, or grains containing large amounts of carbohydrates that are not digested and absorbed by the small bowel. A low carbohydrate diet will almost always markedly reduce flatus volume in patients who do not swallow excessive air.

Most patients with excessive flatus excrete mainly H_2 and CO_2 (Bond and Levitt, unpublished observations). Therefore, if rectal gas analysis is not available, the first approach should be to rule out carbohydrate malabsorption and reduce dietary carbohydrate. If the patient is bothered by excessive flatus only because he is concerned about the possibility of serious underlying organic disease, simple reassurance may be the only therapy required.

Acute Gastric Dilatation

Acute dilatation of the stomach is characterized by rapid, often massive, distention of the stomach and duodenum with gas and fluid, without actual organic obstruction. It may occur following trauma, surgery, or severe illness. Dragstedt & Dragstedt (22) showed that ligation of the distal duodenum in dogs or mechanical obstruction of the pylorus or proximal small bowel did not usually lead to marked gastric dilatation since the normally functioning stomach expelled gas and fluid back through the esophagus. Thus, marked dilatation is secondary to a disturbance of motor function of the stomach and/or gastroesophageal junction in combination with rapid swallowing or aspiration of large quantities of air. Routine use of nasogastric intubation both during and after major operations under general anesthesia has greatly decreased the frequency of acute gastric dilatation in surgery (23). This complication should still be suspected in any patients with trauma or serious illness who develops acute abdominal distention and tympany, and immediate gastric intubation and decompression should be performed.

The "Gas-Bloat" Syndrome Following Surgery

Surgical treatment of gastroesophageal reflux may produce gas-related complaints by interfering with the normal elimination of swallowed air. Currently, the most effective operation, the Nissen fundoplication, involves wrapping a cuff of stomach around the lower esophagus to reinforce a weak sphincter mechanism. Following this operation, approximately 20% of patients are unable to belch and 25–50% note abdominal distention and increased flatus (12). Presumably, the operation causes distention by impairing the ability to eructate swallowed air. Unable to escape by the usual mechanism, most of this air passes on through the bowel and is excreted from the rectum. In most of these patients, the gas-bloat syndrome is transient (24) and is best treated by counseling the patient to avoid habitual aerophagia.

Patients who must resort to esophageal speech after total laryngectomy also may develop a form of the gas-bloat syndrome. In order to speak, these individuals learn to force large volumes of air into the esophagus and then rapidly eructate while articulating words. Some of these patients develop bothersome abdominal distention followed by increased passage of flatus because excessive air is propelled into the stomach and small bowel during extended periods of talking.

Intestinal Gas and Bowel Obstruction

Gaseous distention of bowel increases the likelihood of gut necrosis or perforation following mechanical bowel obstruction or severe ileus. Most of the gas accumulating proximal to a bowel obstruction appears to be swallowed air containing primarily N_2 (25). Wangensteen (26) produced simple obstruction of the terminal ileum in dogs; an average of 328 ml of gas accumulated and the animals lived only about five days. If the cervical esophagus was first interrupted, little gas accumulated and the dogs survived 35 days. Based on these studies, the routine use of nasogastric or small bowel intubation has contributed greatly to the successful management of bowel obstruction or ileus.

The role of gas in the pathophysiology of a closed loop obstruction in man has received little study. Wangensteen found that closed loop obstructions in fasting dogs contained only a few milliliters of gas (26). This observation does not, however, rule out the possibility that the gas which forms in acute closed loop obstructions in man, such as cecal or sigmoid volvulus, could be of bacterial origin if appropriate substrates were present in the segment at the time of obstruction. Compromise of the blood supply to the segment would prevent absorption of the evolved gases and they would accumulate rapidly. If the gas that accumulates in a cecal or sigmoid volvulus is swallowed air, one must postulate some sort of one-way valve mechanism which allows air to pass into, but not out of, the obstructed segment.

Gas Explosion in the Colon

Two of the gases produced in the colon, CH_4 and H_2, are combustible and, in the proper mixtures with O_2, may be explosive (27). There have been many reports of life-threatening explosions resulting from electrocautery ignition of these gases during electrosurgical treatment of colonic lesions through the rigid proctosigmoidoscope (28–30). Although the fiberoptic colonoscope has been used for electrocautery snare resection of colonic polyps in probably over 100,000 patients in the US, no such explosions have been reported during colonoscopy. Presumably, the fasting and the extensive bowel cleansing required for colonoscopy lowers the concentration of H_2 substrates and H_2- and CH_4-producing organisms. Analysis of intraluminal gas samples collected during colonoscopy reveal concentrations of H_2 and CH_4 well below the minimum explosive levels (31, 32).

Special precautions are, however, clearly required to prevent electrocautery explosions in the patient undergoing rigid proctosigmoidoscopy or fiberoptic proctosigmoidoscopy who is not prepared in the vigorous manner required for colonoscopy. In this situation, where hazardous concentrations of both H_2 and CH_4 may be encountered, the risk of explosion may be reduced by first aspirating gas from the lumen and then perfusing the operative site with an inert gas such as CO_2 prior to coagulation.

Literature Cited

1. Hippocrates: *Des vents, Oeuvres Complètes d'Hippocrate.* 1849. Paris. Cited by Kantor, J. L. 1918. A study of atmospheric air in the upper digestive tract. *Am. J. Med. Sci.* 155:829–57
2. Levitt, M. D. 1971. Volume and composition of human intestinal gas. *N. Engl. J. Med.* 284:1394
3. Maddock, W. G., Bell, J. L., Tremaine, M. J. 1949. Gastro-intestinal gas. *Ann. Surg.* 130:512
4. Morris, C. R., Ivy, A. C., Maddock, W. G. 1947. Mechanism of acute abdominal distention. *Arch. Surg.* 55: 101–24
5. Steggerda, F. R. 1968. Gastrointestinal gas following food consumption. *Ann. NY Acad. Sci.* 150:57
6. Murphy, E. L., Calloway, D. H. 1972. The effect of antibiotic drugs on the volume and composition of intestinal gas from beans. *Am. J. Dig. Dis.* 17:639
7. Levitt, M. D., French, P., Donaldson, R. M. 1968. Use of hydrogen and methane excretion in the study of the intestinal flora. *J. Lab. Clin. Med.* 72:988 (Abstr.)
8. Engel, R. R., Levitt, M. D. 1970. Intestinal tract gas formation in newborns (Abstr.). In program for *1970 American Pediatric Society and Society for Pediatric Research,* p. 266
9. Levitt, M. D. 1969. Production and excretion of hydrogen gas in man. *N. Engl. J. Med.* 281:122
10. Levitt, M. D., Ingelfinger, F. J. 1968. The volume, composition, and rate of accumulation of human intestinal gas. *Gastroenterology* 54:1296
11. Bond, J. H., Engel, R. R., Levitt, M. D. 1971. Factors influencing pulmonary methane excretion in man. *J. Exp. Med.* 133:572
12. DeMeester, T. R., Johnson, L. F., Kent, A. H. 1974. Evaluation of current operations for the prevention of gastroesophageal reflux. *Ann. Surg.* 180:511
13. Hood, J. H. 1966. Clinical considerations of intestinal gas. *Ann. Surg.* 163:359
14. Kirk, E. 1949. The quantity and composition of human colonic flatus. *Gastroenterology* 12:782
15. Bond, J. H., Levy, M., Levitt, M. D. 1976. Explosion of hydrogen gas in the colon during proctosigmoidoscopy. *Gastrointest. Endosc.* 23:41–42
16. Levitt, M. D., Berggren, T., Hastings, J., Bond, J. H. 1974. Hydrogen (H_2) catabolism in the colon of the rat. *J. Lab. Clin. Med.* 84:163
17. Levitt, M. D. 1969. Production and excretion of hydrogen gas in man. *N. Engl. J. Med.* 281:122
18. Rune, S. J. 1972. Acid-base parameters of duodenal contents in man. *Gastroenterology* 62:533
19. Lasser, R. B., Bond, J. H., Levitt, M. D. 1975. The role of intestinal gas in functional abdominal pain. *N. Engl. J. Med.* 293:524
20. Lasser, R. B., Levitt, M. D., Bond, J. H. 1976. Studies of intestinal gas after ingestion of a standard meal. *Gastroenterology* 70:906 (Abstr.)
21. Bond, J. H., Levitt, M. D. 1976. Quantitative measurement of lactose absorption. *Gastroenterology* 70:1058
22. Dragstedt, L. R., Dragstedt, C. A. 1922. Acute dilatation of the stomach. *J. Am. Med. Assoc.* 79:612
23. Gillesby, W. J., Wheeler, J. R. 1956. Acute gastric dilation. *Am. Surg.* 22:1154
24. DeMeester, T. R., Johnson, L. F. 1975. Evaluation of the Nissen antireflux procedure by esophageal monometry and twenty-four hour pH monitoring. *Am. J. Surg.* 129:94
25. Hibbard, J. S. 1936. Gaseous distention associated with mechanical obstruction of the intestine. *Arch. Surg.* 33:146
26. Wangensteen, C. H., Rea, C. E. 1939. The distention factor in simple intestinal obstruction. *Surgery* 5:327
27. Levy, E. I. 1954. Explosions during lower bowel electrosurgery. *Am. J. Surg.* 88:754
28. Becker, G. L. 1953. Prevention of gas explosions in the large bowel during electrosurgery. *Surg. Gynecol. Obstet.* 97:463
29. Carter, H. G. 1952. Explosion in the colon during electrodesiccation of polyps. *Am. J. Surg.* 84:514
30. Bond, J. H., Levy, M., Levitt, M. D. 1976. Explosion of hydrogen gas in the colon during proctosigmoidoscopy. *Gastrointest. Endosc.* 23:41
31. Ragins, J., Shinya, H., Wolff, W. I. 1974. The explosive potential of colonic gas during colonoscopic electrosurgical polypectomy. *Surg. Gynecol. Obstet.* 138:554–56
32. Bond, J. H., Levitt, M. D. 1975. Factors affecting the concentration of combustible gases in the colon during colonoscopy. *Gastroenterology* 68:1445–48

Ann. Rev. Med. 1980. 31:139–57
Copyright © 1980 by Annual Reviews Inc. All rights reserved

MYOCARDIAL BIOPSY[1]

◆7349

Chuichi Kawai, M.D. and Akira Matsumori, M.D.

Third Division, Department of Internal Medicine, Faculty of Medicine,
Kyoto University, Kyoto, Japan

Keishiro Kawamura, M.D.

Third Division, Department of Internal Medicine, Osaka Medical College,
Takatsuki, Japan

HISTORICAL REVIEW

Biopsy of the human myocardium has become increasingly important in determining the diagnosis, severity, prognosis, and, in certain instances, the therapy of cardiomyopathies. Surgical biopsies requiring thoracotomy (1) were replaced by punch needle biopsies using modified Vim-Silverman needles either during limited thoracotomy or percutaneously via the transthoracic approach (2–5). Menghini needles, more easily handled and less traumatic to the subject (6), are occasionally used in myocardial biopsy, puncturing the left ventricle to approach the ventricular septum (7), and in pericardial biopsy (8). Potential hazards of these percutaneous needle biopsies are cardiac tamponade and pulmonary complications during the ventricular puncture.

Biopsy of the human myocardium by means of an intravascular catheter would certainly have advantages. General anesthesia or thoracotomy would not be necessary to obtain multiple specimens. Cardiac tamponade or pneumothorax would no longer be a problem. In 1962, Sakakibara & Konno (9) first developed a biopsy catheter or "bioptome" that could obtain biopsy specimens of the endocardium and the myocardium safely and easily. The practicality, safety, and capability of this instrument in providing adequate samples for histologic studies were first confirmed in dogs (10). Subsequently it has been successfully used in human beings (11–14). More recently, Bulloch and associates (15) introduced a catheter-sheathed, long, curved biopsy needle through the right external or internal jugular vein to

---

[1] This study was supported in part by a Research Grant for Cardiomyopathy from the Ministry of Health and Welfare, Tokyo, Japan.

0066-4219/80/0401-0139$01.00

obtain samples from the right ventricular septum. The intracardiac approach is now the most widely accepted method of myocardial biopsy. Reports on the usefulness of Konno's endomyocardial biopsy catheter come from Uganda (16), Austria (17), the United States (18), Sweden (19), England (20), and France (21).

Caves and his associates at Stanford University designed a new biopsy forceps, similar to Konno's, which is introduced percutaneously into the right internal jugular vein and advanced to the right ventricle (22–25). Their original purpose was to diagnose acute rejection in cardiac transplant recipients by serially repeated myocardial biopsies. They extended their technique to the study of patients with cardiomyopathy (26). A similar endomyocardial biopsy forceps for the right and left ventricle had been developed in Europe and used with a guide catheter as a sheath (27).

Although the use of Konno's bioptome has been considered simple and safe, its stiffness complicates intravascular and intracardiac manipulation and cleaning the blood trapped inside the Teflon® sleeve creates difficulties. Modifications of the original bioptome were necessary. Two such modified biopsy catheters are now available. One, the Olympus GFB biopsy catheter (28), closely resembles Konno's bioptome except that there is no Teflon sleeve covering the shaft. Another modification, King's endomyocardial bioptome (29, 30), resembles the Olympus fiberoptic bronchoscope biopsy forceps with a polyolefin covering on the shaft. Hirota et al (31) evaluated two methods of endomyocardial biopsy, the Konno bioptome and the Olympus fiberoptic bronchoscope biopsy forceps in dogs. They concluded that the latter was easier to handle because it was used with a thin-walled Cournand catheter as a guide sheath. It was also smaller and easier to clean.

Brooksby and associates (32, 33) solved both the problem of percutaneous introduction and that of accurate placement of the biopsy catheter with the minimum of manipulation when they developed the St. Thomas' long sheath technique. Left ventricular biopsy, which was difficult with the stiff Konno bioptome, is now possible with either the small-sized Konno bioptome or King's biopsy forceps properly placed via the sheath passed retrogradely across the aortic valve or transseptally across the mitral valve (34–37).

Kawai & Kitaura (38), on the other hand, facilitated left ventricular biopsy by designing a new biopsy catheter with a flexible tip controlled by rotating a knob on the operating handle. This biopsy catheter makes it easier to enter the left ventricle and obtain multiple specimens in succession.

Endomyocardial biopsy of the right and left ventricles in pediatric patients is now being attempted with Konno's pediatric bioptome, King's biopsy forceps, or, even smaller, a modified Machida bronchial biopsy forceps with or without a guide tube of ultrathin radiopaque Teflon (39, 40).

TECHNIQUE

The procedure of transvenous endomyocardial biopsy most commonly used is described here. The Konno bioptome has an ordinary intracardiac catheter shaft 100 cm long, with cutting pincers or jaws at the exploring end and a proximal operating handle. The sliding assembly on the handle is connected to the pincers by a single wire running through the catheter shaft. The pincers, consisting of two mobile cups, open and close by pushing and pulling of the sliding assembly. The diameter of the pincers is 2.5 mm for the smaller and 3.5 mm for the larger Konno bioptomes. The right saphenous vein, routinely used for biopsy of the right atrium and ventricle, is exposed; the bioptome with the pincers closed is introduced into the vein and advanced to the heart under fluoroscopic control. The left antecubital vein may also be used to reach the apex of the right ventricle. The bioptome may sometimes be difficult to pass through the inferior vena cava at its junction with the femoral, renal, and hepatic veins. However, slight rotation of the bioptome will usually allow easy passage.

When the bioptome reaches the right ventricle, contact with the ventricular wall is recognized by a sensation of resistance, the fluoroscopic appearance, and an elevation of the ST segment in the intracardiac electrocardiogram recorded from the exploring end of the bioptome. Still under fluoroscopic control, the bioptome is pulled back slightly, the sliding assembly pushed forward so that the pincers are opened, and the bioptome is readvanced and pressed against the wall of the ventricle. The pincers close as the sliding assembly is pulled back forcefully. The bioptome is immediately withdrawn; one may feel some resistance as a piece of the tissue is excised. At this stage there may be one or two ventricular ectopic beats. The specimen is approximately 2 X 3 mm. The same procedure can be repeated to obtain several specimens. For right atrial biopsy, Konno et al (41) suggested taking a specimen from the external dorsal side or from the septum of the internal dorsal side.

When left ventricular biopsy is attempted, the Konno bioptome is advanced retrogradely via the femoral artery or left brachial artery. However, because of its stiffness, it is sometimes difficult to pass the bioptome through the aortic valve into the left ventricle. The Kawai biopsy catheter obviates this problem. When advance of the biopsy catheter is hampered at the aortic valve, the physician gently bends the flexible tip (Figure 1) enabling the biopsy catheter to pass through the aortic valve into the left ventricle very easily. The cardiac tissue is excised just as with the Konno bioptome except that a straight stylet is advanced through the lumen of the catheter shaft to provide enough stiffness for excision of the tissue by the pincers. The pincers of the Kawai biopsy catheter are similar in size to Konno's; there-

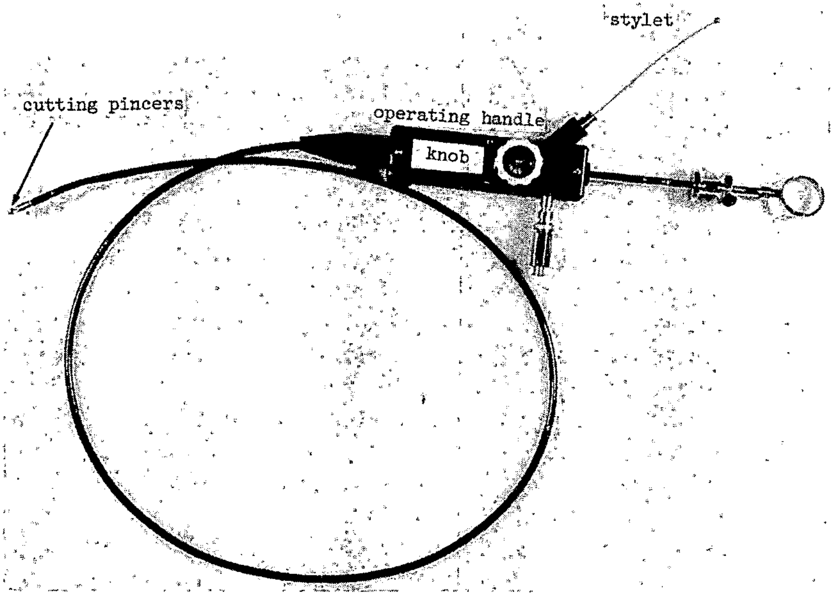


Figure 1 The Kawai flexible endomyocardial biopsy catheter. The catheter tip can be bent about 40° in one direction and about 10° in the opposite direction by rotating a knob on the operating handle.

fore, the specimens obtained by both the Kawai and Konno catheters are larger than those obtained by King's bioptome, which is smaller. Introduction of the biopsy catheter into the femoral artery requires arteriotomy, which is recommended as safer for Japanese patients whose arteries are usually smaller than those of Europeans and Americans. Arteriotomy also allows systemic heparinization, which prevents peripheral embolism during left ventricular biopsy.

COMPLICATIONS

Since the introduction of the transvascular intracardiac approach in 1962 (9), myocardial biopsy has proven surprisingly safe in experienced hands. Unpleasant sensations or pain in the chest occur in a few cases, particularly at the moment of taking the specimen (41, 42), but there have been no subsequent problems. In a large series of more than 1300 right ventricular biopsies by the Stanford group (37), cardiac perforation occurred four times (0.3%), but all four patients recovered uneventfully. There were no deaths. One or two premature ventricular beats are common when a biopsy catheter is pressed against the ventricular wall, but these are harmless provided the biopsy proceeds rapidly. In a series of 188 patients in whom the long sheath technique was used for left ventricular biopsy (32), there were six episodes of systemic embolism (3.2%), all cerebral. In our experience there were no cardiac perforations in approximately 300 right ventricular biopsies with the Konno bioptome. Two patients developed permanent right bundle branch block, probably because the specimens were taken from a position high in the right ventricular septum via a saphenous vein approach. There was no morbidity or mortality in 70 left ventricular biopsies using the Konno or the Kawai biopsy catheter under systemic heparinization. A recent world survey (43) on cardiac biopsy demonstrated cardiac tamponade due to cardiac perforation in 30 (0.5%), perforation of the aortic valve in 1 (0.02%), and death in 3 (0.05%) of 6642 cases. These rates of complications are lower than those in routine cardiac catheterization. Thus, transvascular endomyocardial biopsy can be performed with minimum risk and relative ease.

PATHOLOGY

A minute biopsy specimen may or may not accurately reflect the changes occurring in the rest of the myocardium. The value of the histological analysis of endomyocardial biopsy specimens is controversial: helpful information can be obtained in 70% of patients (44), but in only 10% is diagnostically useful information obtainable (36). However, several investigators (36, 44, 45) found a good correlation between biopsy and autopsy changes. In a few cases, much more advanced interstitial fibrosis was found at autopsy than at the time of biopsy (13). In one series the importance of electron microscopic examination was emphasized as of diagnostic value in endomyocardial biopsy (36). Biopsy specimens certainly have an advantage over necropsy specimens, which show considerable agonal and postmortem autolytic changes. The present review discusses the value of myocardial biopsy in determining the diagnosis, severity, prognosis, and nature of cardiomyopathies.

Hypertrophic Cardiomyopathy

Characteristic histologic features in asymmetric septal hypertrophy (hypertrophic cardiomyopathy) are hypertrophied, bizarrely shaped, and abnormally arranged cardiac muscle cells in the ventricular septum (46). Electron micrographs of muscle cells reveal increased cellular branching, extensive side-to-side intercellular junctions, widened Z bands, and evidence of formation of new sarcomeres (47). These abnormalities are either absent or rare in muscle from the left and right ventricular free walls of patients with the obstructive type of hypertrophic cardiomyopathy. In contrast, numerous disorganized cells are widely distributed in the left and right ventricular free walls in the nonobstructive type (46). It is now known that these changes do not necessarily constitute a pathognomonic sign for the diagnosis of hypertrophic cardiomyopathy. However, biopsy specimens obtained from the left ventricular outflow tract at the time of therapeutic cardiac operation (46, 47), and from the right ventricular septum by endomyocardial bioptome (48), occasionally reveal marked disarray of hypertrophied myocardial cells (Figure 2) and of myofibrils and myofilaments within individual myocytes (Figure 3). These findings may support a presumptive clinical diagnosis of hypertrophic cardiomyopathy. Such foci of myocardial disarray also occur in certain patients with other types of heart disease, but usually to a smaller extent than in hypertrophic cardiomyopathy.

The endomyocardial biopsy specimens are usually too small to evaluate the extent of disarray of myocardial cells. Using autopsy studies (49) on patients with hypertrophic cardiomyopathy, two with and one without obstruction, we quantitatively assessed the diagnostic value of endomyocardial biopsy in this disorder. Two large cross sections (upper and lower one third) of the hearts were each divided into 3 to 5 pieces, sectioned at 4 μm, and stained with hematoxylin-eosin. The distribution of marked myocardial fiber disarray is shown as the hatched area in Figure 4. Lines were drawn at a depth of 2 mm from the endocardial surface of the right ventricular septum and of the left ventricular free wall to demarcate the region of endomyocardial biopsy. A summing up of the sections of the lines crossing the hatched area revealed that the probability of hitting the site of marked myocardial fiber disarray by right ventricular endomyocardial biopsy would be no more than 33% in a patient with obstructive hypertrophic cardiomyopathy.

Our study suggested that biopsy of the right ventricular septum was superior to that of the left ventricular free wall in detecting marked myocardial fiber disarray in obstructive hypertrophic cardiomyopathy. Davies et al (50) found this characteristic histological appearance in left ventricular endomyocardial biopsies of only one of six cases of asymmetric septal hypertrophy. In nonobstructive hypertrophic cardiomyopathy, endomyocardial biopsy from the left ventricular free wall might increase the probability of finding specific changes.

The left ventricle is the chamber mainly involved in many adult patients with cardiomyopathy. Therefore, biopsy specimens from the left ventricle are generally more informative than those from the right (51). This is particularly true in patients with nonobstructive hypertrophic cardiomyopathy, especially in the early asymptomatic stage (51). It is important to correctly diagnose cardiomyopathy as early as possible. The pathological features of left ventricular biopsy specimens differed from those of right ventricular specimens in 60% of 20 patients, some of whom suffered from cardiomyopathy (33). Left ventricular biopsy should be considered espe-

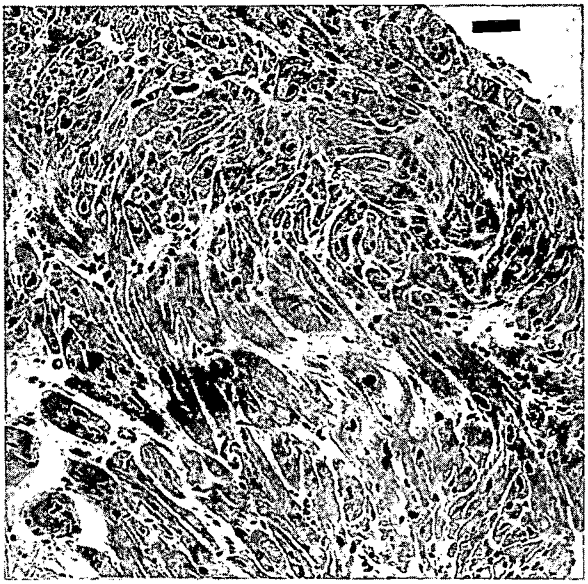


Figure 2 Light microscopy of a biopsied specimen obtained from the right ventricular septum in a patient with hypertrophic nonobstructive cardiomyopathy. The myocardial fibers are hypertrophied, bizarrely shaped, and abnormally arranged, showing a whirl formation. Hematoxylin-eosin stain. Scale bar: 40 μm.

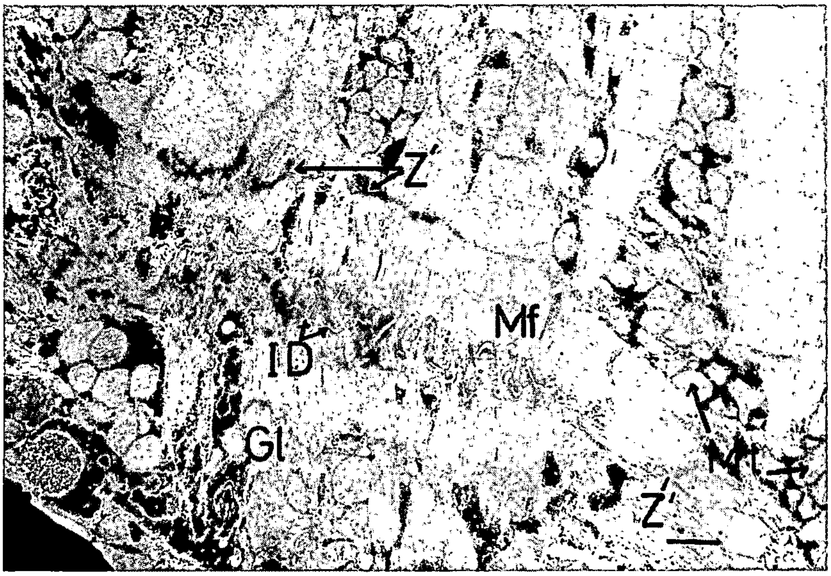


Figure 3 Electron microscopy of a biopsied specimen obtained from the left ventricular free wall in a patient with hypertrophic nonobstructive cardiomyopathy. Marked myofibrillar (Mf) disarray, expansions of Z-band material (Z'), distortion of the intercalated disc (ID), accumulation of glycogen granules (Gl), and mitochondria (Mt). Scale bar: 1 μm

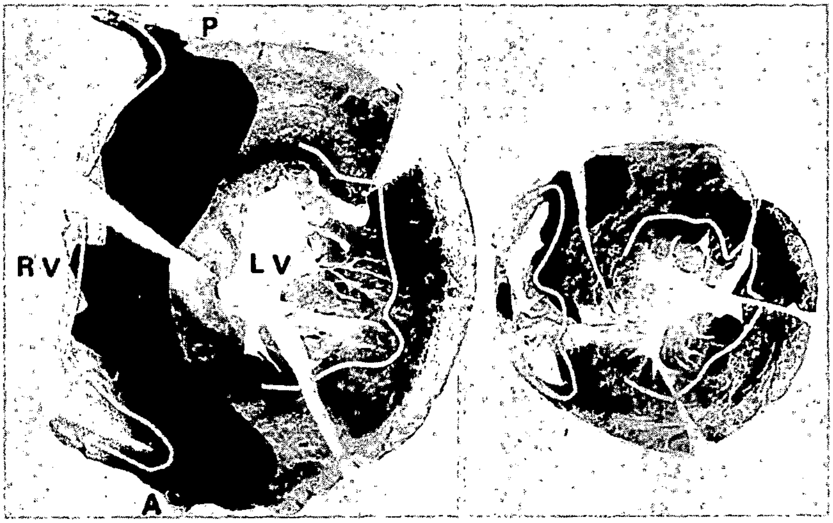


Figure 4 Distribution of marked myocardial fiber disarray in a cross section of the heart: upper one third (*left*) and lower one third (*right*). Dark hatched area is the location of marked disarray of the myocardial fibers in the interventricular septum and the left ventricular free wall. White lines represent a depth of 2 mm from the endocardial surface of the right and left ventricles that can be reached by endomyocardial biopsy. P: posterior, A: anterior, RV: right ventricle, LV: left ventricle.

cially for early diagnosis of nonobstructive hypertrophic cardiomyopathy. In another series, however, of 15 patients mainly with suspected congestive cardiomyopathy, there were no significant differences between right and left ventricular specimens in the majority of cases (52). Diffuse pathological processes such as congestive cardiomyopathy, particularly in the advanced stage, might have shown identical features in the right and left ventricular biopsy specimens. Our own experiences in congestive cardiomyopathy are discussed below.

Determining the intramyocardial catecholamine content in hypertrophic cardiomyopathy in relation to its pathogenesis has been a challenging problem. As the norepinephrine content of the myocardium was far above the normal level in hypertrophic cardiomyopathy, it was called noradrenosis (53). However, this finding could not be confirmed by a recent study from the same laboratory (54). On the other hand, the concentration of norepinephrine in the atrium as well as the left ventricle removed at the time of operation was reduced in patients with heart failure (55, 56).

Specimens obtained during open heart surgery under general anesthesia are clearly unsuitable for the analysis of catecholamines. For that purpose biopsy specimens obtained during cardiac catheterization would be more natural and accurate. We developed a method of determining the myocardial content of catecholamine from a small biopsy specimen obtained during diagnostic cardiac catheterization. Biopsy specimens are immediately frozen in liquid nitrogen. The tissues are homogenized in 100 μl of freshly prepared cold 0.4 N perchloric acid per milligram wet weight of tissue, containing 5 mM reduced glutathione in a glass tissue-grinder submerged in an ice bath, and then centrifuged at 4°C for 15 minutes at 30,000 X g to produce a "protein-free" supernatant. An aliquot of this clear supernatant is injected directly into the column of our newly designed catecholamine-analyzing system using high performance liquid chromatography coupled with fluorometric assay (57). We are now accumulating data from patients with hypertrophic cardiomyopathy or congestive cardiomyopathy with or without heart failure. This is potentially another useful application of myocardial biopsy.

Congestive (Dilated) Cardiomyopathy

Although the biopsy specimens fail to disclose pathognomonic histologic or ultrastructural features that could be used to establish the diagnosis (58, 59), there is generally a good correlation between the severity of the clinical condition and the extent of the morphologic changes (60).

Between 1964 and 1978, 52 right and 23 left ventricular endomyocardial biopsies were obtained from 59 patients with congestive cardiomyopathy (CCM). In the 16 patients who underwent both right and left ventricular biopsies, the biopsies were performed on the same day or 7–10 days apart.

Conventional light microscopy revealed a variety of histological changes in the biopsies. They included disarrangement of myocardial fibers, hypertrophy of myocardial fibers, scarcity of myofibrils, nuclear changes in myocardial fibers, vacuolization and basophilic degeneration of myocardial fibers, deposition of a peculiar substance in the sarcoplasm, proliferation of collagen fibers, interstitial edema, fatty infiltration, cell infiltration, and endocardial thickening. Most of these changes varied markedly in degree so that they were arbitrarily graded 1 to 4+ depending upon the severity and extent. In general, 1+ signified slight change, 2+ moderate change, 3+ marked change, 4+ extreme change, and (–) no apparent abnormality. In assessing hypertrophy of myocardial fibers, transverse diameters of 100 consecutive myofibers crossing through the nuclei were magnified 600 times and measured by a calibrated eye reticule under the light microscope. The mean diameter and its standard deviation were calculated for each biopsy. Mean diameters ranging from 16–20 μm, 20–24 μm, 24–28 μm, and more than 28 μm were graded as 1, 2, 3, and 4+ of hypertrophy, respectively. A mean diameter of less than 16 μm was graded as (–), signifying no hypertrophy of myocardial fibers. Hemorrhage or myofibrillar contracture bands were usually excluded as artifacts. Overall histologic gradings are listed in Table 1.

Table 1 Congestive cardiomyopathy

Histological findings	Right ventricular biopsies (no.)	Gradings					Left ventricular biopsies (no.)	Gradings				
		–	1+	2+	3+	4+		–	1+	2+	3+	4+
Disarrangement of myofibers	52	13 (25)[a]	12 (23)	19 (37)	8 (15)	0	23	7 (30)	7 (30)	6 (26)	3 (14)	0
Hypertrophy of myofibers	48	13 (27)	24 (50)	7 (15)	4 (8)	0	21	0	6 (29)	8 (38)	4 (19)	3 (14)
Scarcity of myofibrils	52	3 (6)	19 (36)	24 (46)	6 (12)	0	23	0	4 (17)	13 (57)	5 (22)	1 (4)
Nuclear changes	52	6 (12)	22 (42)	21 (40)	3 (6)	0	23	1 (4)	6 (27)	12 (53)	2 (8)	2 (8)
Vacuolization	52	22 (42)	24 (46)	5 (10)	0	1 (2)	23	6 (26)	10 (44)	6 (26)	0	1 (4)
Basophile degeneration	52	48 (92)	4 (8)	0	0	0	23	21 (91)	2 (9)	0	0	0
Deposition of peculiar substance in myofibers	52	51 (98)	0	1 (2)	0	0	23	22 (96)	0	1 (4)	0	0
Fibrosis	52	13 (25)	25 (48)	10 (19)	4 (8)	0	23	3 (13)	10 (44)	9 (39)	0	1 (4)
Interstitial edema	52	10 (19)	35 (67)	5 (10)	2 (4)	0	23	8 (35)	11 (48)	4 (17)	0	0
Fatty infiltration	52	34 (65)	11 (21)	5 (10)	1 (2)	1 (2)	23	22 (96)	1 (4)	0	0	0
Cell infiltration	52	51 (98)	1 (2)	0	0	0	23	20 (87)	3 (13)	0	0	0
Endocardial thickening	49	32 (66)	11 (22)	4 (8)	1 (2)	1 (2)	20	13 (65)	1 (5)	4 (20)	2 (10)	0

[a] Figures in parentheses indicate the percentage of incidence.

As shown in Table 1, hypertrophy and nuclear change of myocardial fibers and scarcity of myofibrils were more prominent in the left ventricular endomyocardial biopsies than in the right. In contrast, interstitial edema and fatty infiltration were more common in the right ventricular endomyocardial biopsies. Although not tabulated here, the same tendency was evident in the 16 patients in whom both right and left ventricular biopsies were obtained.

In our observations, almost 60% of either right or left ventricular biopsies revealed a slight (1+) to moderate (2+) grade of myofiber disarray, and almost 15% exhibited marked (4+) fiber disarray (Table 1). The marked fiber disarray appeared to be uncommon, but fiber disarray up to a moderate degree was fairly common in the right ventricular endomyocardial biopsies obtained from patients with various heart conditions excluding cardiomyopathy. Widespread disarray of myocardial fibers has been accepted as a characteristic feature of the ventricular myocardium, especially the interventricular septum, of patients with hypertrophic cardiomyopathy (46, 47, 60, 61). The specificity and significance of this abnormality, however, remain uncertain (62–64). In their studies on ventricular biopsies, Ferrans and his colleagues (60, 61) observed marked disarray of myocardial fibers in 10% of patients with CCM and in 18% of patients with congenital heart diseases associated with right ventricular hypertrophy and obstruction to right ventricular outflow. It should be remembered, therefore, that the disarray of myocardial fibers seen in the ventricular biopsies is not specific for the diagnosis of hypertrophic cardiomyopathy.

In patients with CCM, electron microscopy of ventricular endomyocardial biopsies often revealed a wide spectrum of hypertrophic and/or degenerative changes of the myocardium.

In general, the cardiocytes were connected with one another at end-to-end intercellular junctions (intercalated discs). Also seen were multiple intercalated discs (65–67), widespread side-to-side junctions that were not related to normally oriented intercalated discs, and intracytoplasmic junctions that were formed by the apposition of two areas of the plasma membranes of the same cell (68). These peculiar forms of cell junctions occurred more commonly in the areas of marked disarray of myocardial cells.

The myofibrils appeared to be disoriented near the periphery of hypertrophied myocardial cells. These myofibrils were often associated with abnormal Z band expansions. In degenerated cardiocytes, the myofibrils underwent various degrees of lysis, which were characterized by streaming and fragmentation of Z bands and disorganization and disappearance of myofilaments. A Z-band-like material was often associated with the sarcolemma of the hypertrophied or degenerated cardiocytes.

The mitochondria often varied in size; many were smaller and more numerous (mitochondriosis) than normal. Some contained inclusions of glycogen. In degenerated cardiocytes, inner mitochondrial membranes were often scarce.

The sarcotubules exhibited dilatation and/or proliferation in some cardiocytes. In hypertrophied cardiocytes, rough-surfaced profiles of the endoplasmic reticulum were commonly seen. The transverse tubules often became less prominent and even hard to identify in degenerated cardiocytes.

The nuclei were markedly hypertrophied and bizarrely shaped in a number of cardiocytes. Adjacent to the nuclei, Golgi complexes were often increased in size and lipofuscin granules were common.

The ultrastructural changes of the biopsies from our patients with CCM resembled those reported by others (60, 69, 70), and were probably not specific to the disease. Similar findings were documented for other types of heart diseases of known etiology (71–73).

By conventional light microscopy or transmission electron microscopy, three-dimensional imaging of individual cardiocytes or intercellular connections requires reconstruction from a number of serial sections of the cells, an impractical procedure.

But scanning electron microscopy permits elucidation of the three-dimensional morphology of individual myocardial cells (74) or the architecture of some connected cells. In this new field, a few studies have been conducted on autopsied human hearts with hypertrophic cardiomyopathy (75, 76). In the future, this approach will likely be used for cardiac biopsies from patients with cardiomyopathy or other types of heart diseases. Attempts to identify more specific evidence for diagnosis in biopsy specimens have so far proven unrewarding.

Immunofluorescence studies for viral etiological agents demonstrated positive fluorescence for adenovirus antigen in only one of more than 20 patients examined (77). Coxsackie B_3 virus was isolated from biopsy material in one of 25 patients with congestive cardiomyopathy (78). Generally, however, demonstration of viral particles and positive viral culture in myocardial biopsy specimens have been unsuccessful in patients with congestive cardiomyopathy (79). Nevertheless, these facts do not eliminate the possibility of previous viral infections no longer reflected by the presence of viral antigens in the myocardium as a cause of cardiomyopathy (77, 80). Moreover, nonspecific interstitial fibrosis in biopsy specimens of patients with suspected congestive cardiomyopathy may well represent a sequela of past myocarditis, in view of the fact that serial biopsy studies revealed a very rapid recovery from histologic changes representing acute myocarditis, which left only nonspecific fibrosis with or without cell infiltration in the myocardium, in cases following a benign course (81).

Direct evidence of acute myocarditis itself is rather difficult to obtain by biopsy in clinical cases. It is essential to obtain biopsy specimens in the very early phase of acute myocarditis before the characteristic findings of acute inflammation disappear. Figure 5 demonstrates the histologic appearance of a biopsy specimen from the right ventricle of a 30-year-old Japanese male who had developed Stokes-Adams syndrome due to complete atrioventricular block eight days previously. There is a mild to moderate degree of interstitial round cell infiltration with focal myocardial necrosis consistent with acute myocarditis. The "flu-like syndrome" with moderate fever preceded the Stokes-Adams syndrome by three days. There was a marked elevation of GOT, LDH, and CPK. Sinus rhythm with 1 : 1 conduction was restored within 24 hours. Thereafter, the entire course was uneventful.

It is extremely important to follow a patient like this with definite evidence of myocarditis to see if myocardial damage occurring during the acute phase of the disease remains for the rest of his life and eventually evolves into cardiomyopathy.

Secondary Myocardial Disease

This is a group of diseases for which endomyocardial biopsy can provide direct and specific evidence to aid in diagnosis. This group includes myocar-

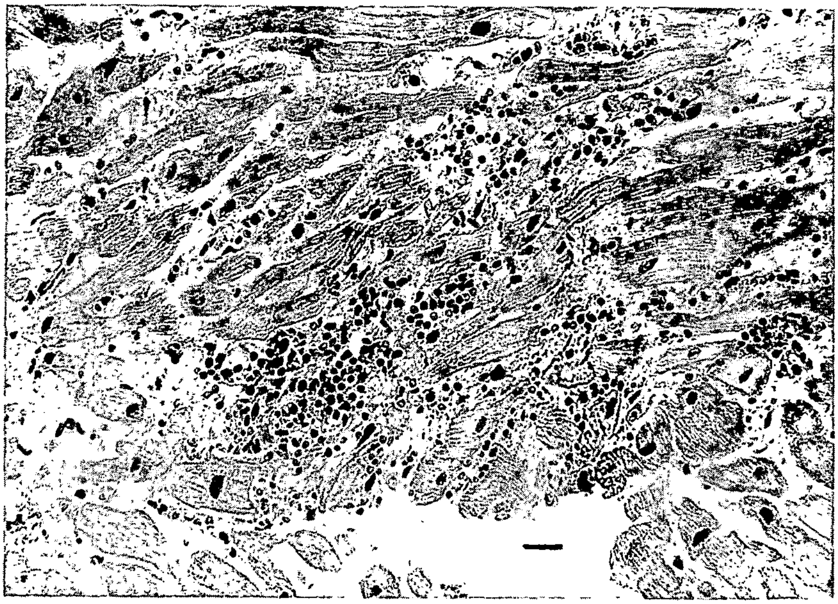
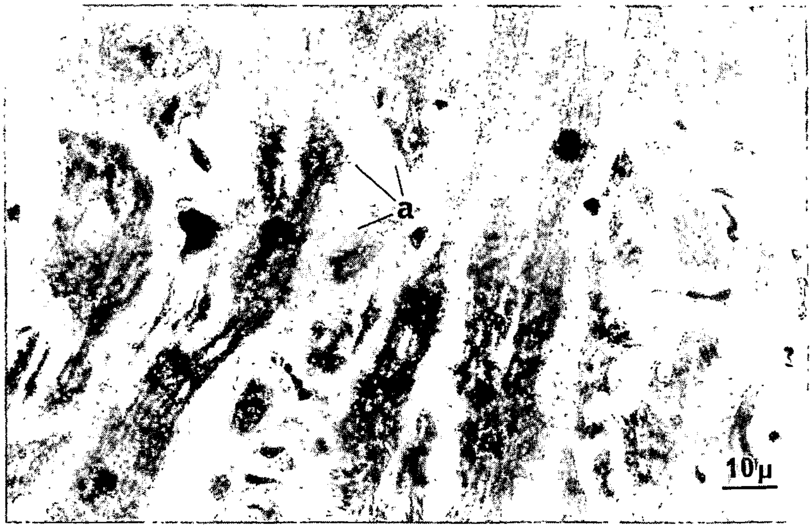


Figure 5 Light microscopy of a biopsied specimen from the right ventricle of a 30-year-old man with acute myocarditis showing interstitial round cell infiltration with myocardial necrosis. Hematoxylin-eosin stain. Scale bar: 20 μm.

dial involvement resulting from infective, infiltrative, metabolic, and toxic diseases. In some cases abnormal deposits of amyloid (Figure 6) iron, calcium, or mucopolysaccharides can be detected with appropriate special stains by light microscopy.

One useful application of endomyocardial biopsy is to detect cardiac amyloidosis, which may otherwise remain unidentified for a long time (82–85). For example, a 60-year-old woman who presented with the sick sinus syndrome, had had transient dizziness with no sequelae three years earlier. One year later, she developed exertional dyspnea and general malaise and was found to have cardiomegaly. Three months prior to admission, she presented with arrhythmias. On admission, her pulse rate was 48 per minute, irregular; blood pressure, 162/72 mm Hg; cardiothoracic ratio, 0.61; no Bence Jones protein was detectable in the urine; ECG revealed a sinus rhythm associated with ectopic atrial rhythm and AV junctional escape rhythm but no low voltage. Right heart catheterization recorded normal pressure curves in the right atrium and ventricle; the mean pressure was also normal in the pulmonary artery and capillaries. An overdrive suppression test supported the diagnosis of sick sinus syndrome. Right ventricular endomyocardial biopsy revealed amorphous material surrounding myocardial fibers and infiltrating into the vascular walls in conventional light microscopic sections (Figure 6, *top*). Stained with Congo red, the peculiar material gave green-yellow birefringence of amyloid under the polarizing microscope. Electron microscopy confirmed numerous fibrils of amyloid adjacent to the surface of cardiocytes and in the intercellular space of the myocardium (Figure 6, *bottom*). Liver and rectal biopsies were also obtained later and amyloid deposit was identified in both. In this particular case of sick sinus syndrome, an early diagnosis of amyloidosis could not have been made without endomyocardial biopsy.

Myocarditis, myocardial sarcoidosis or giant cell myocarditis, and glycogen storage disease (Pompe's disease) can be readily recognized by characteristic histologic findings. Myocardial diseases associated with neurologic and collagen disorders have been suggested by biopsy. Immunofluorescent staining has been used in the detection of immunologically mediated cardiac diseases, such as humoral cardiac transplant rejection, at Stanford University Medical Center (86). Recently the cardiotoxicity of daunorubicin and adriamycin has also been studied by endomyocardial biopsy. Moreover, endomyocardial biopsy has a definite advantage in correctly diagnosing diseases involving the endocardium such as endocardial fibroelastosis, and, sometimes, endomyocardial fibrosis including Davies' disease, endocarditis parietalis fibroplastica (Löffler), and cardiovascular collagenosis (Becker).



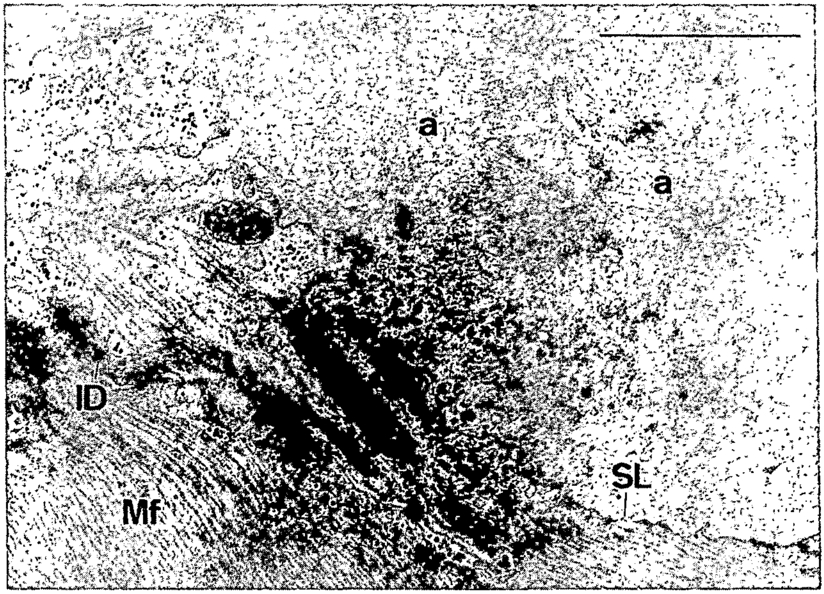


Figure 6 (*top*) Deposit of amyloid substance (a) in right ventricular endomyocardial biopsy. Hematoxylin-eosin stain. Scale bar: 10 μm. (*bottom*) Electron micrograph of a number of amyloid fibrils (a) accumulated in the vicinity of the sarcolemma (SL) of the cardiocyte and in the intercellular space. ID: intercalated disc, Mf: myofibrils. Scale bar: 1 μm.

SUMMARY

The history, technique, complications, and evaluation of myocardial biopsy, particularly in cardiomyopathies, have been reviewed and discussed. As expected, specific histological abnormalities for the diagnosis of cardiomyopathies were hard to obtain in such small fragments of myocardium. Rather, myocardial biopsy at present provides information on the severity and prognosis of cardiomyopathies, especially those with a diffuse disease process. However, morphologic examination is definitely valuable in diagnosing cardiomyopathies, even if it merely confirms the clinical diagnosis, because none of the information presently given by any one diagnostic tool is specific enough to correctly diagnose cardiomyopathy. Endomyocardial biopsy is most useful from a diagnostic standpoint in acute cardiac rejection in transplant recipients as well as in secondary myocardial diseases: myocarditis, amyloidosis, hemosiderosis, glycogen storage disease, sarcoidosis, etc.

Advances in biochemical, immunologic, and more specific morphologic analyses will increase the usefulness of myocardial biopsy in diagnosing and assessing the etiology of idiopathic cardiomyopathy.

ACKNOWLEDGMENT

We are indebted to Drs. K. Ishizawa and H. Fujiwara for preparing the photomicrographs, and to Dr. Alice S. Cary for pertinent advice on the manuscript.

Literature Cited

1. Weinberg, M., Fell, E. H., Lynfield, J. 1958. Diagnostic biopsy of the pericardium and myocardium. *Arch. Surg.* 76:825–29
2. Sutton, D. C., Sutton, G. C., Kent, G. 1956. Needle biopsy of the human ventricular myocardium. *Q. Bull. Northwest. Univ. Med. Sch.* 30:213–14
3. Sutton, D. C., Sutton, G. C. 1960. Needle biopsy of the human ventricular myocardium: Review of 54 consecutive cases. *Am. Heart J.* 60:364–70
4. Raffensperger, J., Driscol, J., Sutton, G. C., Weinberg, M. 1964. Myocardial biopsy. *Arch. Surg.* 89:1021–23
5. Shirey, E. K., Hawk, W. A., Mukerji, D., Effler, D. B. 1972. Percutaneous myocardial biopsy of the left ventricle. Experience in 198 patients. *Circulation* 46:112–22
6. Timmis, G. C., Gordon, S., Baron, R. H., Brough, A. J. 1965. Percutaneous myocardial biopsy. *Am. Heart J.* 70:499–504
7. Bercu, B., Heinz, J., Choudhry, A., Cabrera, P. 1964. Myocardial biopsy. A new technic utilizing the ventricular septum. *Am. J. Cardiol.* 14:675–78
8. Shugoll, G. I. 1973. Percutaneous myocardial and pericardial biopsy with the Menghini needle. *Am. Heart J.* 85:35–39
9. Sakakibara, S., Konno, S. 1962. Endomyocardial biopsy. *Jpn. Heart J.* 3:537–43
10. Wennemark, J. R., Sommers, H. M., Zitnik, R. S. 1966. Endomyocardial biopsy. Experimental study with a catheter technique. *Am. Heart J.* 72:675–80
11. Konno, S., Sakakibara, S. 1963. Endomyocardial biopsy. *Dis. Chest* 44: 345–50
12. Kawai, C., Inoue, K., Yamashita, K., Kurimoto, K., Hanada, H., Takatsu, T. 1966. Studies on endo-myocardial

biopsy. (1) Basic consideration. *Jpn. Circ. J.* 30(Suppl.):72 (Abstr.)

13. Takatsu, T., Kawai, C., Tsutsumi, J., Inoue, K. 1968. A case of idiopathic myocardiopathy with deposits of a peculiar substance in the myocardium; diagnosis by endomyocardial biopsy. *Am. Heart J.* 76:93–104
14. Sekiguchi, M., Konno, S. 1969. Histopathological differentiation employing endomyocardial biopsy in the clinical assessment of primary myocardial disease. *Jpn. Heart J.* 10:30–46
15. Bulloch, R. T., Murphy, M. L., Pearce, M. B. 1965. Intracardiac needle biopsy of the ventricular septum. *Am. J. Cardiol.* 16:227–33
16. Somers, K., Hutt, M. S. R., Pathel, A. K., D'Arbela, P. G. 1971. Endomyocardial biopsy in diagnosis of cardiomyopathies. *Br. Heart J.* 33:822–32
17. Mösslacher, H., Slany, J. 1971. Neue diagnostische Methoden und Geräte. *Z. Kreislaufforsch.* 60:289–91
18. Ali, N., Ferrans, V. J., Roberts, W. C., Massumi, R. A. 1973. Clinical evaluation of transvenous catheter technique for endomyocardial biopsy. *Chest* 63:399–402
19. Torp, A. 1973. Endomyocardial biopsy. *Scand. J. Thorac. Cardiovasc. Surg.* 7:253–61
20. Olsen, E. G. J. 1974. Diagnostic value of the endomyocardial bioptome. *Lancet* 1:658–59
21. Petitier, H., Bouhour, J. B., De Lajartre, A. Y., Crochet, D., Nicolas, G., Horeau, J. 1976. La biopsie endomyocardique. Innocuité et intéret clinique. *Arch. Mal. Coeur Vaiss.* 69:1005–11
22. Caves, P. K., Stinson, E. B., Graham, A. F., Billingham, M. E., Grehl, T. M., Shumway, N. E. 1973. Percutaneous transvenous endomyocardial biopsy. *J. Am. Med. Assoc.* 225:288–91
23. Caves, P. K., Stinson, E. B. 1973. Percutaneous transvenous endomyocardial biopsy in human heart recipients. *Ann. Thorac. Surg.* 16:325–36
24. Caves, P. K., Stinson, E. B., Billingham, M. E., Shumway, N. E. 1974. Serial transvenous biopsy of the transplanted human heart. Improved management of acute rejection episodes. *Lancet* 1:821–26
25. Caves, P. K., Schulz, W. P., Dong, E. Jr., Stinson, E. B., Shumway, N. E. 1974. New instrument for transvenous cardiac biopsy. *Am. J. Cardiol.* 33: 264–67
26. Caves, P., Coltart, J., Billingham, M., Rider, A., Stinson, E. 1975. Transvenous endomyocardial biopsy—application of a method for diagnosing heart disease. *Postgrad. Med. J.* 51:286–90
27. Müller, S., Müller, P., Richter, G. 1971. Transvenöse und transarterielle Endomyokardbiopsie mit einem Katheter-Bioptom. *Z. Kreislaufforsch.* 60:291–303
28. Ali, N. 1974. Transvenous endomyocardial biopsy using the gastrointestinal biopsy (Olympus GFB) catheter. *Am. Heart J.* 87:294–97
29. Richardson, P. J. 1974. King's endomyocardial bioptome. *Lancet* 1:660–61
30. Richardson, P. J. 1975. Technique of endomyocardial biopsy—including a description of a new form of endomyocardial bioptome. *Postgrad. Med. J.* 51:282–85
31. Hirota, Y., Khaja, F., Abelmann, W. H. 1976. Effectiveness and hazard of endomyocardial biopsy in dogs: Comparison of two methods. *Am. Heart J.* 92:767–72
32. Brooksby, I. A. B., Swanton, R. H., Jenkins, B. S., Webb-Peploe, M. M. 1974. Long sheath technique for introduction of catheter tip manometer or endomyocardial bioptome into left or right heart. *Br. Heart J.* 36:908–12
33. Brooksby, I. A. B., Jenkins, B. S., Coltart, D. J., Webb-Peploe, M. M., Davies, M. J. 1974. Left-ventricular endomyocardial biopsy. *Lancet* 2: 1222–25
34. Brooksby, I. A. B., Jenkins, S., Davies, M. J., Swanton, R. H., Coltart, J., Webb-Peploe, M. M. 1977. Left ventricular endomyocardial biopsy I: Description and evaluation of the technique. *Cathet. Cardiovasc. Diagn.* 3:115–21
35. Brooksby, I. A. B., Webb-Peploe, M. M. 1977. Endomyocardial biopsy. *Cathet. Cardiovasc. Diagn.* 3:111–13
36. Mackay, E. H., Littler, W. A., Sleight, P. 1978. Critical assessment of diagnostic value of endomyocardial biopsy. Assessment of cardiac biopsy. *Br. Heart J.* 40:69–78
37. Mason, J. W. 1978. Techniques for right and left ventricular endomyocardial biopsy. *Am. J. Cardiol.* 41:887–92
38. Kawai, C., Kitaura, Y. 1977. New endomyocardial biopsy catheter for the left ventricle. *Am. J. Cardiol.* 40:63–65
39. Mackay, E. H., Pickering, D., Littler, W. A. 1977. Cardiac biopsy in childhood. *Arch. Dis. Child.* 52:785–89
40. Lurie, P. R., Fujita, M., Neustein, H. B. 1978. Transvascular endomyocardial biopsy in infants and small children: de-

scription of a new technique. *Am. J. Cardiol.* 42:453–57

41. Konno, S., Sekiguchi, M., Sakakibara, S. 1971. Catheter biopsy of the heart. *Radiol. Clin. North Am.* 9:491–510
42. Brooksby, I. A. B., Coltart, D. J., Webb-Peploe, M. M. 1975. Progress in endomyocardial biopsy. *Mod. Concepts Cardiovasc. Dis.* 44:65–69
43. Sekiguchi, M., Take, M. 1979. World survey of catheter biopsy of the heart. In *Cardiomyopathy: Clinical, Pathological and Theoretical Aspects,* ed. M. Sekiguchi, E. G. V. Olsen, Tokyo: Tokyo Univ. Press. In press
44. Olsen, E. G. J. 1977. Endomyocardial biopsies: is pathological examination of use? *Br. Heart J.* 39:349 (Abstr.)
45. Sekiguchi, M., Konno, S. 1971. Diagnosis and classification of primary myocardial disease with the aid of endomyocardial biopsy. *Jpn. Circ. J.* 35:737–54
46. Maron, B. J., Ferrans, V. J., Henry, W. L., Clark, C. E., Redwood, D. R., Roberts, W. C., Morrow, A. G., Epstein, S. E. 1974. Differences in distribution of myocardial abnormalities in patients with obstructive and nonobstructive asymmetric septal hypertrophy (ASH). Light and electron microscopic findings. *Circulation* 50:436–46
47. Ferrans, V. J., Morrow, A. G., Roberts, W. C. 1972. Myocardial ultrastructure in idiopathic hypertrophic subaortic stenosis. A study of operatively excised left ventricular outflow tract muscle in 14 patients. *Circulation* 45:769–92
48. Alexander, C. S., Gobel, F. L. 1974. Diagnosis of idiopathic hypertrophic subaortic stenosis by right ventricular septal biopsy. *Am. J. Cardiol.* 34: 142–51
49. Fujiwara, H., Kawai, C. 1978. Clinicopathological studies on myocardial fiber disarray. Part II; Probability of detection of marked disarray in hypertrophic cardiomyopathy by endomyocardial biopsy. *Proc. 46th Kinki Reg. Meet. Jpn. Circ. Soc.*
50. Davies, M. J., Brooksby, I. A. B., Jenkins, S., Cankovic-Darracott, S., Swanton, R. H., Coltart, D. J., Webb-Peploe, M. M. 1977. Left ventricular endomyocardial biopsy II: The value of light microscopy. *Cathet. Cardiovasc. Diagn.* 3:123–30
51. Kawai, C., Takatsu, T. 1975. Clinical and experimental studies on cardiomyopathy. *N. Engl. J. Med.* 293: 592–97
52. Richardson, P. J., Olsen, E. G. L., Jewitt, D. E., Oram, S. 1975. Percutaneous technique of left ventricular biopsy, and comparison between right and left ventricular myocardial samples. *Br. Heart J.* 37:556 (Abstr.)
53. Pearse, A. G. E. 1964. The histochemistry and electron microscopy of obstructive cardiomyopathy. In *Cardiomyopathies,* ed. G. E. W. Wolstenholme, M. O'Connor. pp. 132–64. London: Churchill. 428 pp.
54. Noorden, S. V., Olsen, E. G. J., Pearse, A. G. E. 1971. Hypertrophic obstructive cardiomyopathy, a histological, histochemical, and ultrastructural study of biopsy material. *Cardiovasc. Res.* 5:118–31
55. Chidsey, C. A., Braunwald, E., Morrow, A. G., Mason, D. T. 1963. Myocardial norepinephrine concentration in man. Effects of reserpine and of congestive heart failure. *N. Engl. J. Med.* 269:653–58
56. Chidsey, C. A., Braunwald, E., Morrow, A. G. 1965. Catecholamine excretion and cardiac stores of norepinephrine in congestive heart failure. *Am. J. Med.* 39:442–51
57. Yui, Y., Kimura, M., Itokawa, Y., Kawai, C. 1979. Ultramicro method for the determination of picogram amounts of norepinephrine and epinephrine by high performance liquid chromatography. *J. Chromatogr.* 177:376–79
58. Davies, M. J., Kennedy, S., Brooksby, I. A. B., Darracott, S. C., Jenkins, B. S., Braimbridge, M. V., Coltart, D. J., Webb-Peploe, M. M. 1977. Left ventricular endomyocardial biopsy III: ultrastructural characteristics of cardiomyopathy and cardiac hypertrophy with good or poor ventricular function. *Cathet. Cardiovasc. Diagn.* 3:131–37
59. Olsen, E. G. J. 1978. Endomyocardial biopsy (Editorial). *Br. Heart J.* 40: 95–98
60. Ferrans, V. J., Massumi, R. A., Shugoll, G. I., Ali, N., Roberts, W. C. 1973. Ultrastructural studies of myocardial biopsies in 45 patients with obstructive or congestive cardiomyopathy. *Recent Adv. Stud. Cardiac Struct. Metab.* 2:231–72
61. Ferrans, V. J. 1978. Myocardial ultrastructure in human cardiac hypertrophy. In *Cardiomyopathy and Myocardial Biopsy,* ed.M. Kaltenbach, F. Loogen, E. G. J. Olsen, pp. 100–20. Berlin: Springer. 337 pp.
62. Beçu, L., Sommerville, J., Gallo, A. 1976. "Isolated" pulmonary valve stenosis as part of more widespread car-

diovascular disease. *Br. Heart J.* 38:472–82
63. Bulkley, B. H., Weisfeldt, M. L., Hutchins, G. M. 1977. Asymmetric septal hypertrophy and myocardial fiber disarray. *Circulation* 56:292–98
64. van der Bel-Kahn, J. 1977. Muscle fiber disarray in common heart disease. *Am. J. Cardiol.* 40:355–64
65. Laks, M. M., Morady, F., Adomian G. E., Swan, H. J. C. 1970. Presence of widened and multiple intercalated discs in the hypertrophied canine heart. *Circ. Res.* 27:391–401
66. Maron, B. J., Ferrans, V. J. 1973. Significance of multiple intercalated discs in hypertrophied human myocardium. *Am. J. Pathol.* 73:81–97
67. Adomian, G., Laks, M. M., Morady, F., Swan, H. J. C. 1974. Significance of the multiple intercalated disc in the hypertrophied canine heart. *J. Mol. Cell. Cardiol.* 6:105–9
68. Buja, L. M., Ferrans, V. J., Maron, B. J. 1978. Intracytoplasmic junctions in cardiac muscle cells. *Am. J. Pathol.* 74:613–48
69. Knieriem, H. J. 1978. Electron-microscopic findings in congestive cardiomyopathy. See Ref. 61, pp. 71–86
70. Kunkel, B., Lapp, H., Kober, G., Kaltenbach, M. 1978. Ultrastructural evaluations in early and advanced congestive cardiomyopathies. See Ref. 61, pp. 87–99
71. Maron, B. J., Ferrans, V. J., White, R. I. Jr. 1973. Unusual evolution of acquired infundibular stenosis in patients with ventricular septal defect. Clinical and morphologic observations. *Circulation* 48:1092–1103
72. Maron, B. J., Ferrans, V. J., Roberts, W. C. 1975. Myocardial ultrastructure in patients with chronic aortic valve disease. *Am. J. Cardiol.* 35:725–39
73. Maron, B. J., Ferrans, V. J., Roberts, W. C. 1975. Ultrastructural features of degenerated cardiac muscle cells in patients with cardiac hypertrophy. *Am. J. Pathol.* 79:387–434
74. Powell, T., Steen, E. M., Twist, V. W., Woolf, N. 1978. Surface characteristics of cells isolated from adult rat myocardium. *J. Mol. Cell. Cardiol.* 10:287–92
75. Tulusan, A. H., Brandt, G. 1977. Morphologic differential diagnosis of idiopathic cardiomyopathies. *Virchows Arch. A Pathol. Anat. Histol.* 374:71–78
76. Kawamura, K., Uehara, H., Noda, S., Takatsu, T. 1979. Scanning electron microscope study of normal or hypertrophied myocardial cells isolated from autopsied human hearts. See Ref. 43
77. Kawai, C. 1971. Idiopathic cardiomyopathy. A study on the infectious-immune theory as a cause of the disease. *Jpn. Circ. J.* 35:765–70
78. Kuhn, H., Breithardt, G., Knieriem, H.-J., Loogen, F., Both, A., Schmidt, W. A. K., Stroobandt, R., Gleichmann, U. 1975. Die Bedeutung der endomyokardialen Katheterbiopsie für die Diagnostik und die Beurteilung der Prognose der kongestiven Kardiomyopathie. *Dtsch. Med. Wochenschr.* 100:717–23
79. Ferrans, V. J., Roberts, W. C. 1978. Myocardial biopsy: A useful diagnostic procedure or only a research tool? *Am. J. Cardiol.* 41:965–67
80. Kawai, C., Matsumori, A., Kitaura, Y., Takatsu, T. 1978. Viruses and the heart: viral myocarditis and cardiomyopathy. *Prog. Cardiol.* 7:141–62
81. Sekiguchi, M., Hiroe, M., Morimoto, S., Kawagoe, Y. 1978. Contribution of endomyocardial biopsy in the diagnosis and assessment of cardiomyopathies. *Proc. 8th World Congr. Cardiol., Tokyo.* In press
82. Hedner, P., Rausing, A., Steen, K., Torp, A. 1975. Diagnosis of cardiac amyloidosis by myocardial biopsy. *Acta Med. Scand.* 198:525–28
83. Schroeder, J. S., Billingham, M. E., Rider, A. K. 1975. Cardiac amyloidosis. Diagnosis by transvenous endomyocardial biopsy. *Am. J. Med.* 59:269–73
84. Chew, C., Ziady, G. M., Raphael, M. J., Oakley, C. A. 1975. The functional defect in amyloid heart disease. The "stiff heart" syndrome. *Am. J. Cardiol.* 36:438–44
85. Swanton, R. H., Brooksby, I. A. B., Davies, M. J., Coltart, D. J., Jenkins, B. S., Webb-Peploe, M. M. 1977. Systolic and diastolic ventricular function in cardiac amyloidosis. Studies in six cases diagnosed with endomyocardial biopsy. *Am. J. Cardiol.* 39:658–64
86. Mason, J. W., Billingham, M. E., Rider, A. K., Harrison, D. C. 1977. Myocardial biopsy. In *Clinical Cardiology,* ed. J. T. Willerson, C. A. Sanders, Vol. III of *Science and Practice of Clinical Medicine,* ed. J. P. Sanford, pp. 606–13. New York: Grune & Stratton. 660 pp.

Ann. Rev. Med. 1980. 31:159–66
Copyright © 1980 by Annual Reviews Inc. All rights reserved

TREATMENT OF CHRONIC MYELOID LEUKEMIA

❖7350

P. A. Stryckmans, M.D.

Section of Hematology, Service de Medecine, Institut Jules Bordet,
Brussels, Belgium

INTRODUCTION

Acute lymphoblastic leukemia (ALL) can now be cured in a large porportion of children (1). A cure for acute myeloblastic leukemia, (AML) may have been achieved recently for a small fraction of adult patients (2). Chronic myeloid leukemia (CML) has almost the same prognosis it had 50 years ago, but recent success in eradicating the Ph' positive cell clones holds promise of a cure for CML. This review considers only the Ph' positive CML, by far the most frequent type.

CHRONIC PHASE OF CML

Controlling Growth of Ph' Positive Cells

Many agents were tested only for their capacity to induce remission in the chronic phase of CML, i.e. to alleviate most abnormalities except the Ph' chromosome. Nitrogen mustard, vinblastine, thioguanine, colcemid, chlorambucil, 6-mercaptopurine, cyclophosphamide, etc (3), all given as monochemotherapy,. can induce remission. However, in a randomized trial busulfan was found superior to the last three of these drugs (4–6).

Only busulfan (7, 8), hydroxyurea (9, 10), dibromomannitol (11, 12), and melphalan (13) were evaluated in single-drug regimens during the entire chronic phase. None of them significantly lengthens the survival over that obtained by Minot in 1924 with supportive treatment (14) (Table 1). For busulfan, the response rate in untreated CML is around 98%. It is usually administered for several weeks at a daily dose which should not exceed 4–6 mg and should be stopped when the WBC reaches 10–20,000/μl (7, 8).

0066-4219/80/0401-0159$01.00

Table 1 Results of selected therapies for chronic myeloid leukemia

Treatment	Number of patients	All have Ph'+	Median survival (months)	Mode of treatment[a]	Survival[b]	Ref.	Year
Supportive	52	no	36[c]	—	S	14	1924
Busulfan daily	30	no	42	D	S	7	1961
Busulfan single doses	17	yes	32	D	D	15	1977
Busulfan daily[f]	48	no	40	C	T	31	1968
Splenic irradiation[f]	54	no	28	D	T	31	1968
Dibromomannitol	73	no	43	C	D	11	1973
Hydroxyurea	43	yes	39.6	C	—	10	1973
Busulfan[f]	20	yes	45	D	D	12	1975
Dibromomannitol[f]	20	yes	45	D	D	12	1975
Melphalan	32	yes[d]	28.6	C	T	13	1978
L-5 protocol[e]	37	yes	50	D	D	37	1979

[a] D: discontinuous, C: continuous.

[b] S; from first symptoms, D: from diagnosis, T: from beginning of treatment.

[c] Mean survival.

[d] Except one Ph'– case.

[e] See text.

[f] Randomized trial.

When overdosage is expected, it might be more convenient to give a single dose of 50–150 mg every four weeks (15). After induction of remission whether busulfan should be given continuously at a low maintenance dose or discontinued until the WBC reaches 50–100,000/μl has never been clearly answered in terms of patient survival. The existence of several drugs enables switching to another drug when toxicity or resistance (without evidence of transformation) appears for one agent. Whether it is preferable to use one of these drugs first and the other thereafter in a certain order is not known. There is some indication that the sequences dibromomannitol-busulfan (12) and busulfan-hydroxyurea are useful.

Polychemotherapy of CML has not yet been completely evaluated. A cyclical combination of seven drugs each used for an eight-week period is presently being evaluated (cited in 16). Recently, seven patients were treated with busulfan (2 mg), 6-mercaptopurine (50 mg), and allopurinol (300 mg) daily (17). The mean time to achieve a normal leucocyte count was only 4.5 weeks, and the total dose of busulfan at that point was only one sixth of that used when busulfan was given as a single agent.

Extracorporeal radiation of the blood (18), repeated leukapheresis (19), and immunotherapy (20) require further evaluation.

Treating the Spleen

Early splenectomy has been proposed as an adjuvant therapy to delay blast crisis, based on observations that the spleen may be the source of blastic transformation (21–23). Whether splenectomy is in fact effective in delaying blastic transformation cannot be unequivocally determined from five of the published studies (10, 24–27) because they did not include random nonsplenectomized control groups. The Medical Research Council in Great Britain and the Italian Cooperative Study Group on CML are presently conducting randomized trials on early splenectomy in CML. The most recent analysis of the last study indicates that splenectomized patients do not survive longer than nonsplenectomized patients (M. Baccharani, personal communication 1979). On the other hand, splenectomy for decreased platelet production from long-term busulfan therapy appears useful since it gave an immediate and sustained rise in platelets in 5 of 6 patients (28). Postoperative death after splenectomy during the chronic phase now averages less than 1%, but infectious and thromboembolic postoperative morbidity remains high (29).

Late splenectomy performed in the terminal phase of CML for splenic infarction, persistent abdominal discomfort due to massive splenomegaly, thrombocytopenia, and anemia yielded significant improvement in half of the patients and caused only 4% mortality when performed by an experienced team (30).

Radiotherapy over the spleen was compared with busulfan administered at a maximum daily dose of 4 mg followed by a maintenance dose given almost indefinitely (31). This study showed that, for quality of remission and length of survival, busulfan was significantly superior to splenic radiation. Splenic radiation as the sole form of treatment of CML has since been abandoned.

A new treatment for enlarged spleens is splenic artery infusion of cytosine arabinoside. It reduces spleen size as well as relieves symptoms in CML patients in the blastic phase (32).

Eradicating the Ph' Positive Cell Population

Eradication of the Ph' positive cells, and possibly cure of CML, has been approached by vigorous chemotherapy with or without radiotherapy during the chronic phase of CML. This approach assumes (*a*) the Ph'+ clone can be eradicated; (*b*) the Ph'+ CML patient has still some Ph'– hemopoietic stem cells; and (*c*) the Ph'+ cells are leukemic but the Ph'– cells are normal.

The first assumption appears reasonable: the Seattle group noted the disappearance of Ph'+ cells in four CML patients given high doses of dimethyl busulfan, cyclophosphamide, and total body radiation followed by the infusion of marrow from a genetically identical twin (33). A rapid hematologic recovery took place in the four patients who were clinically and hematologically normal 22, 23, 26, and 31 months after transplantation. All the metaphases examined then were Ph'– whereas all were Ph'+ before transplantation.

The second assumption, that Ph'− stem cells persist and can recover, is based on many arguments: (*a*) the coexistence of Ph'− and Ph'+ metaphases in the marrow of untreated CML patients (34); (*b*) the coexistence of Ph'− and Ph'+ cell colonies in CML marrow grown in vitro (35); (*c*) the appearance of Ph'− cells as a consequence of unexpected busulfan-induced aplasia (36); and (*d*) the appearance of Ph'− cells as a consequence of intensive chemotherapy (37, 38). The L-5 protocol (37), administered to 37 patients, began with radiation of the spleen followed by splenectomy. Chemotherapy consisting of cytosine arabinoside and thioguanine (and L-asparaginase) was started thereafter. In the responders, a maintenance therapy (hydroxyurea or multiple drug regimen) was administered. Of 29 evaluable cases, 12 had a decreased percentage of Ph'+ cells ranging from 0 to 27% of the baseline value. The reduction of the Ph'+ chromosome was difficult to maintain. It lasted from less than 2 months (3 patients) to more than 3 years (2 patients). The median survival of those patients with reduced Ph'+ cells probably will be appreciably longer than the 34-month median of nonresponders. However, these results should be interpreted cautiously since at diagnosis signs of good prognosis were more frequent among the responders than among the nonresponders. More recently, the same group using a more aggressive protocol (L-15) obtained 57% complete remission with complete disappearance of the Ph'+ chromosome (38). Two patients relapsed at 3 and 4 months, but four remain in remission from 1 to 16 months.

The third assumption, that the Ph'− hemopoietic cells are nonleukemic, is more hazardous. As a matter of fact, experimental studies using G-6PD isoenzymes as cell markers, suggest that the Ph' chromosome could be only a late and secondary event in leukemogenesis since Ph'− lymphocytes (39), and Ph'− granulomonocytic colony-forming cells (40) as well, appear to belong to the leukemic clone. Whatever the true nature of the Ph'− cells, their presence originally (34) or after treatment (36, 37) appears to improve the prognosis of CML.

BLASTIC TRANSFORMATION OF CML

Between 56 and 86% of CML patients expire in blast crisis. In a literature survey from 1969 to 1976, the rate of complete remission on 238 patients treated by all sorts of chemotherapy was only 13% (41). However, survival of the responders was eight months and of the nonresponders only two months (41), so treatment is justified.

An important step has been the recognition that CML blast crisis is heterogeneous with respect to cellular expression and that some characteristics of leukemic blasts have predictive value for response to treatment. The most frequent type of acute leukemia transformation is myeloblastic (42) but 20–30% of the patients have leukemic cells with the morphologic (43), cytochemical, and immunologic (44) characteristics of lymphoblasts. High levels of the enzyme terminal deoxynucleotidyl transferase (a prelymphoid cell marker) in the leukemic cells was observed in 13 of 22 patients with Ph'+ blastic CML, but this had no significant correlation with myeloblastic or lymphoblastic blast cell morphology (45).

ALL-type therapy (daily prednisone and weekly vincristine), presently one of the most effective treatments for transformed CML, induces complete remission in 16–37% of all cases (46, 47). Responsiveness was correlated with hypodiploidy, lymphoblastic morphology (48), and reactivity with antiserum specific for ALL of non-T, non-B type (49). The lymphoblastic morphology was associated with a significantly longer median survival (six months) than was the myeloblastic type (three months). The presence of terminal transferase-positive cells appeared as a better predictor of responsiveness to this treatment since all eight complete remissions belonged to the group of 13 terminal transferase-positive patients (45).

AML-type chemotherapy has been extremely disappointing in transformed CML. A chemotherapy regimen called TRAMPCO(L) (50) deserves more attention. It consists of seven drugs (thioguanine, daunomycin, cytosine arabinoside, methotrexate, prednisone, cyclophosphamide, vincristine, and L-asparaginase) given monthly in moderate doses. It gave a good response in 8 of 19 patients.

High dose chemotherapy, usually with total body radiation and followed by marrow reconstitution, has been tested in a few centers (33, 51–53). The antileukemic treatment differed, but in most cases it consisted of cyclophosphamide (60 mg/kg/day for 2 days) followed by total body radiation. One or several of the following drugs were also given: dimethyl myleran, daunomycin, cytosine arabinoside, 6-thioguanine, epidophyllotoxin, hydroxyurea, etc. For marrow rescue, different sorts of hemopoietic cells have been transplanted: marrow cells from a genetically identical twin (33), allogeneic marrow cells from an HLA-identical sibling (51), autologous marrow cells harvested at diagnosis and stored in liquid nitrogen until blastic transformation (52), and autologous white blood cells stored as above (53). The two modalities using autologous cells can, of course, only control the blast crisis since the procedure implies the reestablishment of the chronic phase. A few observations are already available from these four studies: (*a*) clearance of blast cells from the marrow was obtained in most cases; (*b*) early recurrence of blasts was frequent; (*c*) a few promising long-lasting remissions have been obtained; (*d*) engraftment was rarely a major problem; (*e*) for autotransplantation, chronic phase CML cells from the blood may be more appropriate than those from the marrow; and (*f*) in the study using allogeneic marrow cells, the median survival was only 43 days because of a very high incidence of septicemia and interstitial pneumonia (53).

SUMMARY

Most known chemotherapy agents, used singly, may induce remission of CML. Those that best alleviate symptoms (anemia, splenomegaly, etc) do not change either the Ph' chromosome anomaly in the hemopoietic cells nor the median delay of 3–4 years before fatal blastic transformation.

A new goal has recently been formulated in CML therapy, namely complete remission, i.e. disappearance of all signs of leukemia including the Ph' abnormality. This can be achieved in certain cases by the use of aggressive (radio)chemotherapy during the chronic phase of CML. Whether or not it means cure of CML awaits further investigation. Encouraging results with these new therapeutic modalities during the chronic phase, and the possibility of rescuing toxic marrow aplasia by autologous cryopreserved blood stem cells, make it less and less defensible to await CML transformation before using aggressive "eradication" therapy.

Literature Cited

1. Holland, J. F., Glidewell, O. 1972. Oncologist's reply: survival expectancy in acute lymphocytic leukemia. *N. Engl. J. Med.* 287:769–77
2. Keating, M. J., Bodey, G. P., McCredie, K. B., Freireich, E. J. 1979. Five years survival and remission duration in adult acute myelogenous leukemia. *Proc. 15th Ann. Meet. Am. Soc. Clin. Oncol. Abstr. C-517*
3. Stryckmans, P. A. 1974. Current concepts in chronic myelogenous leukemia. *Semin. Hematol.* 11:101–27
4. Southern Cancer Chemotherapy Cooperative Study Group. 1959. Comparison of chlorambucil and myleran in chronic lymphocytic and granulocytic leukemia. *Am. J. Med.* 27:424
5. Southern Cancer Chemotherapy Cooperative Study Group. Huguley, Ch. M., Grizzle, J., Rundles, R. W. et al. 1963. Comparison of 6-mercaptopurine and busulfan in chronic granulocytic leukemia. *Blood* 21:89–100
6. Kaung, D. T., Close, H. P., Whittington, R. M., Patno, M. E. 1971. Comparison of busulfan and cyclophosphamide in the treatment of chronic myelocytic leukemia. *Cancer* 27:608–12
7. Haut, A., Abbott, W. S., Wintrobe, H. M., Cartwright, G. E. 1961. Busulfan in the treatment of chronic myelocytic leukemia. The effect of long intermittent therapy. *Blood* 17:1–19
8. Galton, D. A. G. 1969. Chemotherapy of chronic myelocytic leukemia. *Semin. Hematol.* 6:323–43
9. Kennedy, B. J. 1972. Hydroxyurea therapy in chronic myelogenous leukemia. *Cancer* 29:1052–56
10. Schwarzenberg, L., Mathé, G., Pouillart, P., Weiner, R., Locour, J., Genin, J., Schneider, M., De Vassal, F., Hayat, M., Amiel, J. L., Schlumberger, J., Jasmin, C., Rosenfeld, C. 1973. Hydroxyurea, leukapheresis and splenectomy in chronic myeloid leukemia at the problastic phase. *Br. Med. J.* 1:700–3
11. Dibromomannitol Cooperative Study Group. 1973. Survival of chronic myeloid leukemia patients treated by Dibromomannitol. *Eur. J. Cancer* 9: 583–89
12. Canellos, G. P., Young, R. C., Nieman, P. E., De Vita, V. T. Jr. 1975. Dibromomannitol in the treatment of chronic granulocytic leukemia: a prospective randomized comparison with busulfan. *Blood* 45:197–203
13. Hauch, Th., Logue, G., Laszlo, J., Cox, E., Rundles, W. 1978. Treatment of chronic granulocytic leukemia with melphalan. *Blood* 51:571–77
14. Minot, G. R., Buckman, T. E., Isaacs, R. 1924. Chronic myelogenous leukemia: duration and benefit derived. *J. Am. Med. Assoc.* 82:1486–94
15. Sullivan, J. R., Hurley, T. H., Bolton, J. H. 1977. Treatment of chronic myeloid leukemia with repeated single doses of busulfan. *Cancer Treat. Rep.* 61:43–45
16. Spiers, A. S. D. 1976. The treatment of chronic granulocytic leukaemia. *Br. J. Haematol.* 32:291–98

17. Allan, N. C., Duvall, E., Stockdill, G. 1978. Combination chemotherapy for chronic granulocytic leukaemia. *Lancet* 2:523
18. Schiffer, L. M., Chanana, A. D., Cronkite, E. P., Greenberg, M. L., Joel, D. D., Schnappauf, H., Stryckmans, P. A. 1966. Extracorporeal irradiation of the blood. *Semin. Hematol.* 3:154–67
19. Vallejos, C. S., McCredie, K. B., Brittin, G. M., Freireich, E. J. 1973. Biological effect of repeated leukapheresis of patients with chronic myelogenous leukemia. *Blood* 42:925–33
20. Sokal, J. E., Aungst, C. W., Synderman, M., Gomez, G. 1977. Immunotherapy of chronic myelocytic leukemia: effects of different vaccination schedules. *Clin. Hematol.* 6(1):129–39
21. Jacquillat, C. R., Chastang, C. R., Tanzer, J., Briere, J., Weil, M., Pereira-Neto, M., Genon-Auclerc, N. F., Schaison, G., Domingo, A., Boiron, M., Bernard, J. 1975. Facteurs de pronostic de la leucémie myéloïde chronique: à propos de 798 observations. *Nouv. Rev. Fr. Hématol.* 15:229–40
22. Mitelman, F., Brandt, L., Nilsson, P. G. 1974. Cytogenetic evidence for spleen origin of blastic transformation in chronic myeloid leukaemia. *Scand. J. Haematol.* 13:87–92
23. Muller-Bérat, C. N., Wantzin, G. L., Philip, P., Baccarani, M., Killmann, S. A. 1977. Agar culture studies of bone marrow, blood, spleen and liver in chronic myeloid leukemia. *Leukemia Res.* 1:123–31
24. Spiers, A. S. D., Baikie, A. G., Galton, D. A. G., Richards, H. G. H., Wiltshaw, E., Goldman, J. M., Catovsky, D., Spencer, J., Peto, R. 1975. Chronic granulocytic leukaemia: effect of elective splenectomy on the course of the disease. *Br. Med. J.* 1:175–79
25. Italian Cooperative Study Group on Chronic Myeloid Leukemia. 1975. A clinical trial of early splenectomy, hydroxyurea and cyclic arabinosyl cytosine, vincristine and prednisone in chronic myeloid leukemia. *Ser. Haematol.* 8,4:121–42 .
26. Ihde, D. C., Canellos, G. P., Schwartz, J. H., De Vita, V. T. 1976. Splenectomy in the chronic phase of chronic granulocytic leukemia. Effects in 32 patients. *Ann. Intern. Med.* 84:17–21
27. McBride, C. M., Hester, J. P. 1977. Chronic myelogenous leukemia: management of splenectomy in high-risk population. *Cancer* 39:653–58
28. Canellos, G. P., Whang-Peng, J., Carbone, P. P. 1972. Splenectomy for thrombocytopenia in chronic granulocytic leukemia. *Cancer* 29:660–65
29. Wolf, D. J., Silver, R. T., Coleman, M. 1978. Splenectomy in chronic myeloid leukemia. *Ann. Intern. Med.* 89:684–89
30. Gomez, G. A., Sokal, J. E., Mittelman, A., Aungst, C. W. 1976. Splenectomy for palliation of chronic myelocytic leukemia. *Am. J. Med.* 61:14–22
31. Medical Research Council's Working Party for Therapeutic Trials in Leukemia. 1968. Chronic granulocytic leukaemia: comparison of radiotherapy and busulfan therapy. *Br. Med. J.* 1:201–8
32. Canellos, G. P., Sutliffe, S. B., De Vita, V. T., Lister, T. A. 1979. Treatment of refractory splenomegaly in myeloproliferative disease by splenic artery infusion. *Blood* 53:1014–17
33. Fefer, A., Cheever, M. A., Thomas, E. D., Boyd, Ch., Ramberg, R., Gluckberg, H., Buckner, C. D., Storb, R. 1979. Disappearance of Ph'-positive cells in four patients with chronic granulocytic leukemia after chemotherapy, irradiation and marrow transplantation from an identical twin. *N. Engl. J. Med.* 300:333–37
34. Sakurai, M., Hayata, I., Sandberg, A. A. 1976. Prognostic value of chromosomal findings in Ph'-positive chronic myelocytic leukemia. *Cancer Res.* 36:313–18
35. Chervenick, P. A., Ellis, L. D., Pan, S. F., Lawson, A. L. 1971. Human leukemic cells: in vitro growth of colonies containing the Philadelphia (Ph') chromosome. *Science* 174:1134–36
36. Galton, D. A. G. 1972. Radical therapy of chronic granulocytic leukemia. Thyroid tumours, lymphomas granulocytic leukemia. In *Proc. Padua Semin. Clin. Oncol., Oct. 1972,* ed. M. Fiorentino, R. Vangelista, E. Grigoletto, pp. 95–97. Padua: Piccin Med. Books
37. Cunningham, I., Gee, T., Dowling, M., Chaganti, R., Bailey, R., Hpfan, S., Bowden, L., Turnbull, A., Knapper, W., Clarkson, B. 1979. Results of treatment of Ph'+ chronic myelogenous leukemia with an intensive regimen (L-5 Protocol). *Blood* 53:375–95
38. Arlin, Z., Kemplin, S., Gee, T., Goto, T., Chaganti, R., Koenisberg, E., Clarkson, B. 1979. Induction of true remission in Philadelphia chromosome positive (Ph'+) chronic myelogenous leukemia (CML) with the L-15 protocol. *Proc. 15th Ann. Meet. Am. Soc. Clin.*

Oncol. May 14–15, Abstr. C-198, p. 339
39. Fialkow, P. J., Denman, A. M., Jacobson, R. J., Lowenthal, M. N. 1978. Chronic myelocytic leukemia: origin of some lymphocytes from leukemic stem cells. *J. Clin. Invest.* 62:815–23
40. Singer, J. W., Fialkow, P. J., Steinmann, L., Najfeld, V., Stein, S. J., Robinson, W. A. 1979. Chronic myelocytic leukemia (CML). Failure to detect residual normal committed stem cells in vitro. *Blood* 53:264–68
41. Canellos, G. P. 1977. The treatment of chronic granulocytic leukaemia. *Clin. Haematol.* 6:113–28
42. Hammouda, F., Quaglino, D., Hayhoe, F. J. G. 1964. Blastic crisis in chronic granulocytic leukemia. Cytochemical, cytogenetic and autoradiographic studies in four cases. *Br. Med. J.* 1:1275–81
43. Boggs, D. R. 1974. Hematopoietic stem cell theory in relation to possible lymphoblastic conversion of chronic myeloid leukemia. *Blood* 44:449–53
44. Janossy, G., Greaves, M. F., Revez, V. T., Lister, T. A., Roberts, M., Durrant, J., Kirk, B., Catovsky, D., Beard, M. E. J. 1976. Blast crisis of chronic myeloid leukemia (CML). II cell surface marker analysis of "lymphoid" and myeloid cases. *Br. J. Haematol.* 34:179–92
45. Marks, S. M., Baltimore, D., McCafrey, R. 1978. Terminal transferase as a predictor of initial responsiveness to vincristine and prednisone in blastic chronic myelogenous leukemia. *N. Engl. J. Med.* 298:812–14
46. Canellos, G. P., De Vita, V. T., Whang-Peng, J., Chabner, B. A., Schein, P. S., Young, R. C. 1976. Chemotherapy of the blastic phase of chronic granulocytic leukemia: hypodiploidy and response to therapy. *Blood* 47:1003–9
47. Marmont, A. M., Damasio, E. E. 1973. The treatment of terminal metamorphosis of chronic granulocytic leukaemia with corticosteroids and vincristine. *Acta Haematol.* 50:1–8
48. Rosenthal, S., Canellos, G. P., Whang-Peng, J., Gralnick, H. R. 1977. Blast crisis of chronic granulocytic leukemia. *Am. J. Med.* 63:542–47
49. Janossy, G., Roberts, M., Greaves, M. F., Woodruff, R., Pippard, M., Prentice, G., Hoffbrand, A. V. 1978. Lymphoid blast crisis in chronic myeloid leukaemia and Philadelphia positive acute lymphoid leukemia. *Boll. Ist. Sieroter. Milan.* 57:355–59
50. Spiers, A. S. D., Goldman, J. M., Catovsky, D., Costello, C., Buskard, N. A., Galton, D. A. G. 1977. Multiple-drug chemotherapy for acute leukemia. The TRAMPCOL regimen: results in 86 patients. *Cancer* 40:20–29
51. Doney, K., Buckner, C. D., Sale, G. E., Ramberg, R., Boyd, C., Thomas, E. D. 1978. Treatment of chronic granulocytic leukemia by chemotherapy, total body irradiation and allogeneic bone marrow transplantation. *Exp. Hematol.* 6:738–47
52. Buckner, C. D., Stewart, P., Clift, R. A., et al. 1978. Treatment of blastic transformation of chronic granulocytic leukemia by chemotherapy, total body irradiation and infusion of cryopreserved autologous marrow. *Exp. Hematol.* 6:96–109
53. Goldman, J. M., Catovsky, D., Hows, J., Spiers, A. S. D., Galton, D. A. G. 1979. Cryopreserved peripheral blood cells functioning as autografts in patients with chronic granulocytic leukaemia in transformation. *Br. Med. J.* 1:1310–13

Ann. Rev. Med. 1980. 31:167–79
Copyright © 1980 by Annual Reviews Inc. All rights reserved

PLASMA EXCHANGE IN GLOMERULONEPHRITIS AND RELATED VASCULITIDES

❖7351

C. M. Lockwood, M.D. and D. K. Peters, M.D.

Royal Postgraduate Medical School, Hammersmith Hospital, London, England

INTRODUCTION

The past several decades have witnessed striking advances in the understanding of allergic disease with identification of various types of allergic tissue injury and definition of mediator mechanisms. In the case of glomerular disease, Dixon and his colleagues described two classes of injury, the first due to antibodies to glomerular basement membrane and the second to immune complexes (1). Against the background of the explosive growth in immunology further progress in understanding of human glomerular disease has been relatively slow: little is yet known of the initiating mechanisms responsible for autoantibody production or of the nature of the antigens that provoke the formation of antigen antibody complexes. Advances in treatment have been even less spectacular. Except for the treatment by steroids and cyclophosphamide of the minimal-change nephrotic syndrome (a condition in which the allergic mechanisms are still obscure), there is little evidence for therapeutic value of any of the currently available immunosuppressive drugs in human nephritis. However, in experimental models of allergic glomerular injury it has been possible to show that various manipulations can modify disease activity: reduction of antibody formation by immunosuppression (2); interference with the humoral inflammatory mediators such as complement (3) or coagulation factors (4); or depletion of polymorphonuclear leucocytes (3, 5). But in man steroids, cytotoxic drugs, and anticoagulants have had little success.

Against this background we developed an approach to the treatment of severe glomerular injury by using intensive plasma exchange to deplete

0066-4219/80/0401-0167$01.00

circulating nephrotoxic macromolecules—autoantibodies or immune complexes—as well as to deplete certain of the inflammatory mediators. This paper describes our experience of this therapeutic regimen over a period of approximately four years.

EFFECT OF PLASMA EXCHANGE ON PLASMA PROTEINS

We have carried out plasma exchange on a Haemonetics Model 30 cell-separator. We found that daily four-liter exchanges could be carried out easily, with little discomfort to the patient, and within 1½ to 2 hours using an arteriovenous Scribner shunt. However, according to necessity, we have also obtained vascular access by subcutaneous arterovenous fistula or by an inferior vena cava catheter introduced by Seldinger's technique through the femoral vein. If a long period of daily plasma exchange was not anticipated, then conventional access using the antecubital veins was found satisfactory. The optimal volume of plasma to be removed at each exchange was determined by using a dye (Evans Blue) as a marker of the intravascular constituents and ^{131}I-labeled IgG (given 48 hours before exchange) as an immunoglobulin marker. A four-liter exchange removed approximately 90% of the Evans Blue and 65% of the previously administered radiolabeled IgG.

The effect of such exchanges, carried out daily, on native plasma components will depend on the constitution of the replacement material and the synthesis rates of the components. The kinetics of synthesis, in turn, may be modified by various factors such as drug therapy and, for the broad class of "acute-phase" reactants, the degree of accompanying inflammatory stimulus. If plasma protein fraction (PPF), a semipure albumin solution, is used for plasma replacement, then in the absence of an inflammatory acute-phase stimulus (as evidenced by normal concentrations of C-reactive protein) daily four-liter plasma exchanges result in a state of profound hypogammaglobulinemia (IgG levels 0.2–0.3 gram per liter), hypocomplementemia (C3 20–40 mg per 100 ml), and hypofibrinogenemia (fibrinogen 80–100 mg per 100 ml).

PLASMA EXCHANGE IN THE MANAGEMENT OF GLOMERULONEPHRITIS

Rapidly progressive ("crescentic" or "malignant") nephritis is an important cause of renal failure. The hallmark of this form of nephritis is the rapid development of renal failure due to an acute necrotizing glomerulitis with extensive crescent formation. Two pathogenetic mechanisms have been identified (1). The first involves the production of antibody to the glomerular basement membrane (GBM): fixation of this in the kidney, and probably also to cross-reacting antigens in the lung basement membrane, produces the combination of nephritis and lung hemorrhage known as Goodpasture's syndrome (6). Anti-GBM antibody can be detected in the circulation of patients with Goodpasture's syndrome by radioimmune assay (7), and its pathogenetic significance has been established by experiments in which antibody eluted from diseased kidneys caused a severe nephritis in subhuman primates (8). The forms of rapidly progressive nephritis not mediated by anti-GBM antibody are generally assumed to be mediated by immune complex (IC), though here the evidence for the pathogenetic role of IC's is less direct. Circulating immune complexes and glomerular deposits of immunoglobulin and complement occur to a variable extent. This type of nephritis is usually part of a multisystem disease such as the microscopic form of polyarteritis, Wegener's granulomatosis, or Henoch-Schönlein purpura; rarely is disease confined to the kidney alone (9).

MONITORING OF PLASMA-EXCHANGE THERAPY IN GLOMERULONEPHRITIS

Immunological Tests

We have used two sensitive radioimmune assays (7, 10) to measure circulating anti-GBM antibodies, which have been detectable in all patients with characteristic renal biopsy immunofluorescence. In immune complex disease, the assays we used—the C1q binding assay and the conglutinin binding assay—were only positive in approximately half of the patients. However, serial measurements in patients showing such circulating immune complexes (CIC) indicated that the tests were reasonably well correlated with disease activity (11).

Target Organ Damage

Kidney damage has been assessed by standard clinical methods—particularly serial examinations of urine sediment, creatinine clearance, and in some patients serial renal biopsies. In anti-GBM disease a reliable index of pulmonary involvement was required. Ewan et al (12) showed that the sequential measurement of kCO was a sensitive test of fresh lung bleeding; daily measurements being superior to serial chest X rays. In immune complex disease, multisystem involvement, presumably due to widespread vasculitic lesions, is a common feature, and a method for monitoring cerebral blood flow and metabolism using the isotope ^{15}O has been applied to our patients (13).

Splenic Function

In patients with IC disease, infrequent small-volume plasma exchange could produce effects greater than could be accounted for on the basis of the simple physical removal of complexes. This was attributed to alteration in the phagocytic activity of the reticuloendothelial system (RES) (14, 15). We tested this possibility using two techniques to measure splenic function in vivo (16). Both involved measurement of disappearance times of autologous labeled erythrocytes. In the first, the patient's red cells were coated with a non-complement-fixing anti-D IgG. The clearance of such cells is by way of Fc receptors on splenic macrophages but such a test is restricted to Rh-D-positive patients. In the second, the patient's red cells were thermally damaged (49.5°C for 20 minutes); normally such cells are rapidly removed by the spleen.

RESULTS OF PLASMA EXCHANGE IN ANTI-GBM DISEASE

We have treated 28 patients with anti-GBM disease: 26 showed evidence of nephritis and lung hemorrhage (by clinical history, chest X ray, or changes in kCO) and two had nephritis alone. The therapeutic regimen has been described elsewhere (17). In general, daily four-liter plasma exchanges were combined with cyclophosphamide 3 mg/kg/day and prednisolone 60 mg/day. More recently azathioprine in a dose of 1 mg/kg/day has been used in addition. The dose of cyclophosphamide was reduced to 2 mg/kg/day and azathioprine was omitted in patients older than 55 years to reduce the risk of infection. Cytotoxic drugs were withdrawn if infection or leucopenia (WCC 4000 mm^{-3}) occurred, and only exceptionally was cyclophosphamide continued for longer than 8 weeks.

Effect on Renal Function

Ten of the 28 patients mentioned above had established anuria by the time of referral and extensive destruction of glomerular architecture was shown by renal biopsy. No recovery of renal function occurred in any patient in this group. Five patients had severe impairment of renal function at presentation (four adults with creatinines > 600 μmol/liter and one four-year-old child with creatinine > 400 μmol/liter). Of these five, deterioration in renal function continued despite plasma-exchange treatment in one; was arrested but later deteriorated in a second; and in a third, gradually improved after a short period of dialysis; a fourth patient died within 72 hours of fulminating septicemia and pulmonary hemorrhage; and in the four-year-old child technical difficulties with vascular access precluded more than three plasma exchanges. Thirteen patients presented with deteriorating renal function (creatinines < 600 μmol/liter), and in 12 of these patients renal function improved and stabilized following institution of plasma-exchange and drug therapy (in two, prior treatment with high dosage immunosuppressive drugs and steroids had failed to reverse the decline in renal function). In the remaining patient, in whom early effective exchange was not possible because of problems with vascular access, rapid and irreversible deterioration in renal function occurred and showed no response to the eventual introduction of plasma exchange.

Control of Lung Hemorrhage

Episodes of lung hemorrhage were rapidly controlled in all but one patient, who died of septicemia, lung abscesses, and pulmonary bleeding 72 hours after referral. A "late" death from lung hemorrhage also occurred in another patient, who had been transferred for maintenance dialysis at another hospital.

Effect of Regimen on Antibody Levels

Although overall antibody levels fell with plasma-exchange treatment (Figure 1), the rate of fall was variable. In some patients (Figure 2), autoanti-

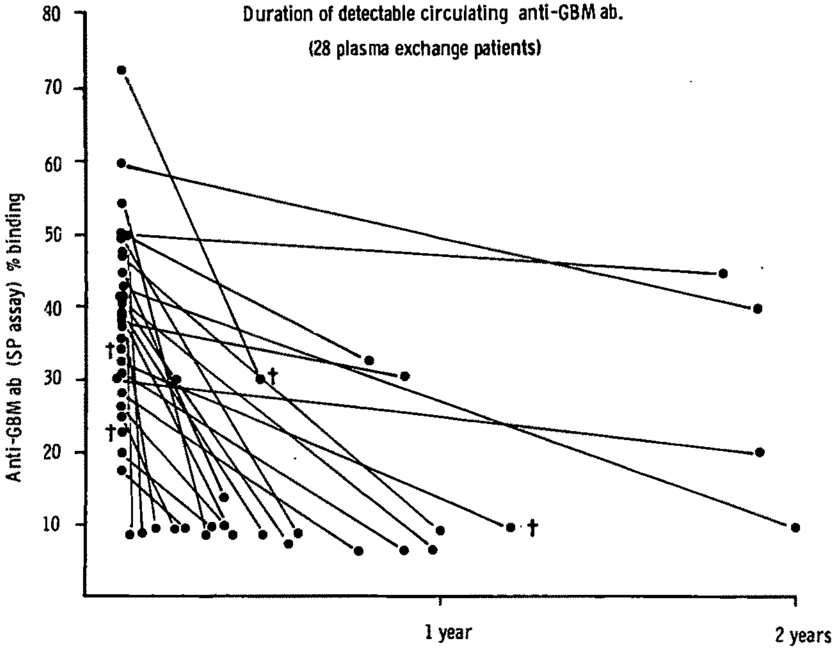


Figure 1 Duration of detectable circulating anti-GBM antibody in 28 plasma-exchange patients as measured by solid phase anti-GBM assay (10). Normal range is less than 10% binding.

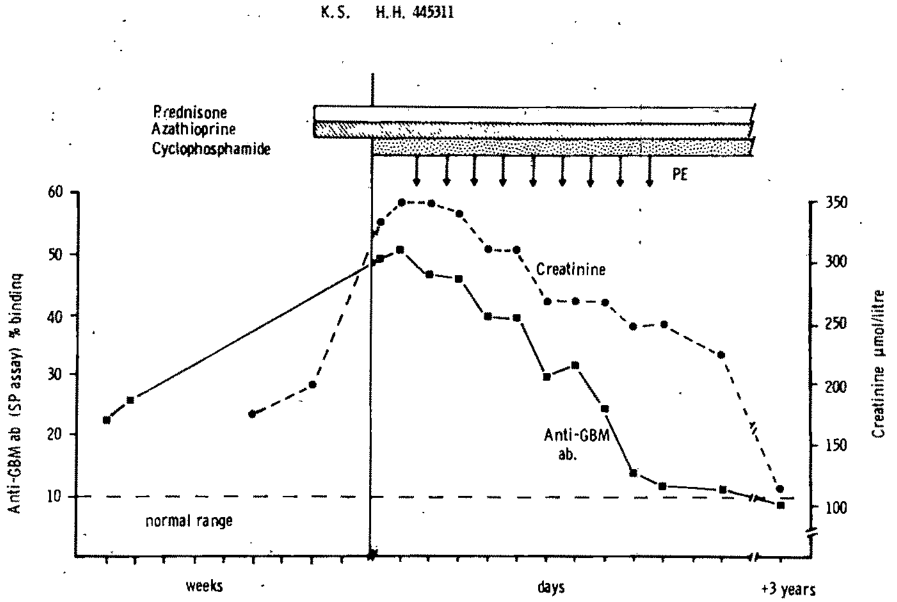


Figure 2 Anti-GBM nephritis. Sequential measurement of serum creatinine and anti-GBM antibody in patient KS. Note that introduction of immunosuppression does not halt rise of either creatinine or antibody.

body synthesis was apparently stopped by the treatment regimen and in these patients the effect was permanent. Improvement in renal function occurred as antibody disappeared from the circulation. In other patients, however, improvement in renal function was maintained despite persistence of circulating antibody. In patients with established anuria plasma exchange was used simply to control lung hemorrhage. Here only a slow fall in antibody levels was observed; however, in this group plasma exchange was generally used for short periods only.

Transplantation

Of four cases studied to date, successful transplantation was carried out only when antibody was no longer detectable (one patient). Transplantation was carried out elsewhere in three patients with circulating antibody. All three transplants failed, and immunofluorescence showed linear deposition of anti-GBM antibody in two: the extent to which recurrent nephritis, as distinct from rejection, contributed to failure is uncertain. The remaining patient suffered hyperacute rejection within 30 minutes, and the kidney showed no linear staining.

Deaths

Seven of the 28 patients have died. In three, death could be ascribed to disease activity—as mentioned above two patients died of lung hemorrhage and a four-year-old child was not offered dialysis support. In the remaining four, infection caused death in two (pneumonia and "hepatitis" six weeks and one year after discharge respectively), aplastic anemia terminating in septicemia occurred in the third after three months, and a myocardial infarct one week after transplant occurred in the fourth.

RESULTS OF PLASMA EXCHANGE IN IC NEPHRITIS

Nineteen patients with a putative IC nephritis have been treated by the plasma-exchange regimen.

Clinical Effects

Improvement in renal function occurred in all 17 patients who were not anuric at referral; the remaining two (anuric) patients showed no recovery of renal function. Improvement in renal disease was closely related to the institution of plasma exchange in 15 of the 17 patients; in two patients renal function improved 2–6 weeks after stopping a series of exchanges. As well as renal function other manifestations of system disease improved concomitantly.

Effects on Immune Complexes

Immune complexes were detectable in 10 of 18 sera available for study at the time of presentation. In addition, cryoglobulins (found in four patients) were present in two more patients in whom tests for immune complexes were negative. The greatest benefit appeared to be in those patients in whom circulating immune complexes had been identified at the start of treatment (Figure 3), though good responses were also seen in CIC-negative patients. In general, detectable circulating immune complexes have not been a feature of follow-up in these patients. In only two patients (on one occasion each) of the 11 patients followed-up for a total of seven "patient-years" were IC's demonstrated.

Deaths

Six patients have died. Three patients died (at 8, 10, and 24 weeks after discharge) because of opportunistic infection, one died from a cerebrovascular accident, and one patient died while on maintenance dialysis therapy; in the remaining patient death was attributed in part to poorly controlled disease activity and in part to a supervening infection.

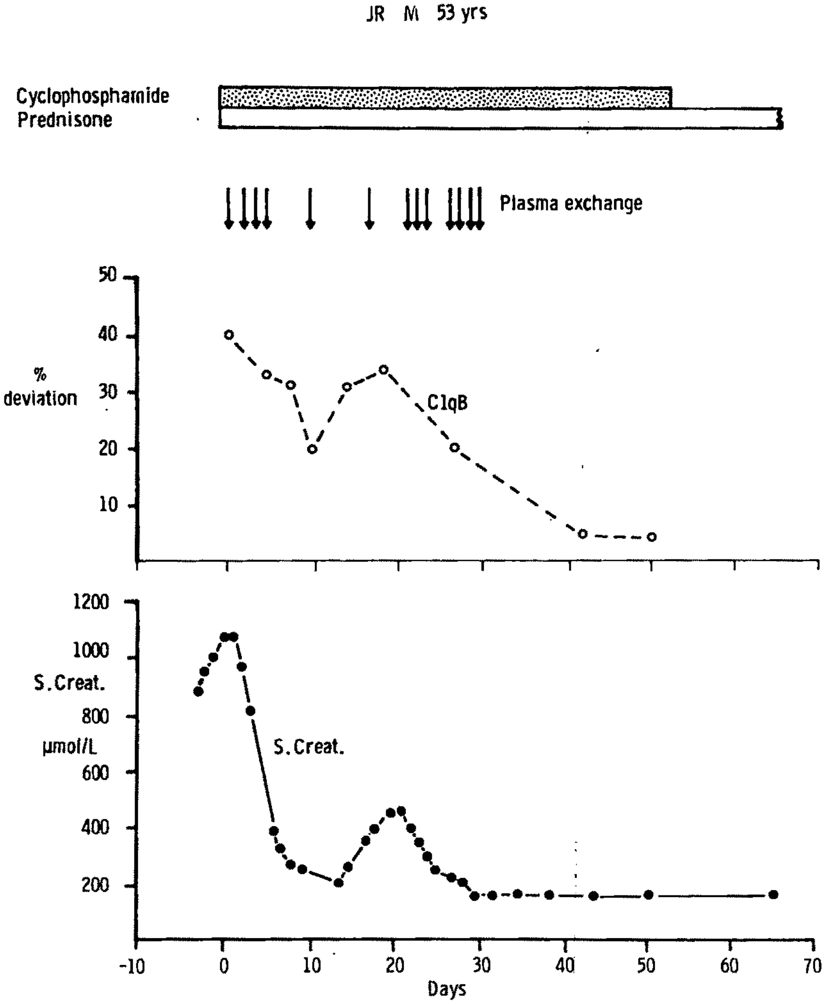


Figure 3 IC nephritis. Serial measurement of serum creatinine and immune complexes as measured by Clq assay (normal < 15% deviation of Clq) in patient JR with Wegener's granulomatosis. Note a second rise in Clq deviation precedes rise in serum creatinine with return to predeterioration values after a second series of plasma exchanges.

Effect on Reticuloendothelial System

Using autologous isotope-labeled IgG-coated or heat-damaged erythrocytes that are normally removed by the spleen, we showed that clearance times for these cells were abnormal in 13 patients with nephritis and/or vasculitis, and that where circulating immune complexes were demonstrated, clearance times were generally correlated with levels of circulating complexes.

Clearance times reverted to normal after a period of plasma exchange in 11 of 13 patients (see Figure 4). In certain patients we showed that levels of complexes continued to fall after plasma exchange was discontinued, possibly as a result of unblocking the impaired phagocytic function brought about by the overwhelming level of immune complexes present before exchange (Figure 5).

DISCUSSION

Plasma-Exchange in other Immunological Diseases

Control of active disease by reducing circulating autoantibody has now been applied to a variety of conditions. They include myasthenia (18), pemphigus (19, 20), rhesus disease (21), and hypoimmunoglobulinemia due to antibodies to B lymphocytes (22). In a significant proportion of patients with anti-GBM disease, the regimen resulted in cessation of anti-GBM antibody synthesis, which suggests that a number of autoantibody-mediated diseases may be cured rather than just controlled by this approach.

Removal of circulating immune complexes with concomitant resolution of disease activity has been reported in various disorders. In systemic lupus

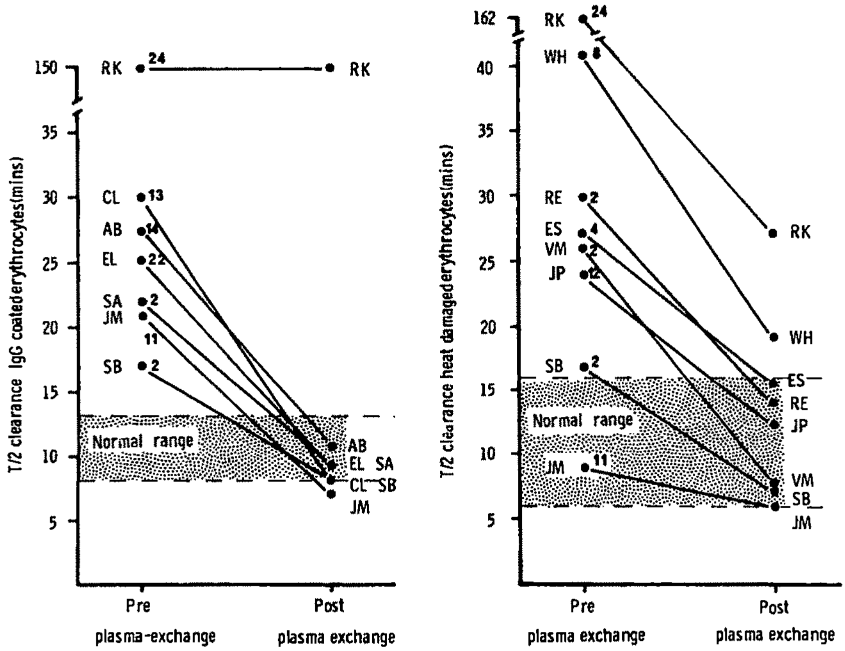


Figure 4 Pre- and post-plasma-exchange measurements of isotope-labeled autologous red cells altered either by thermal damage or coating with IgG. The number of exchanges between measurements is shown.

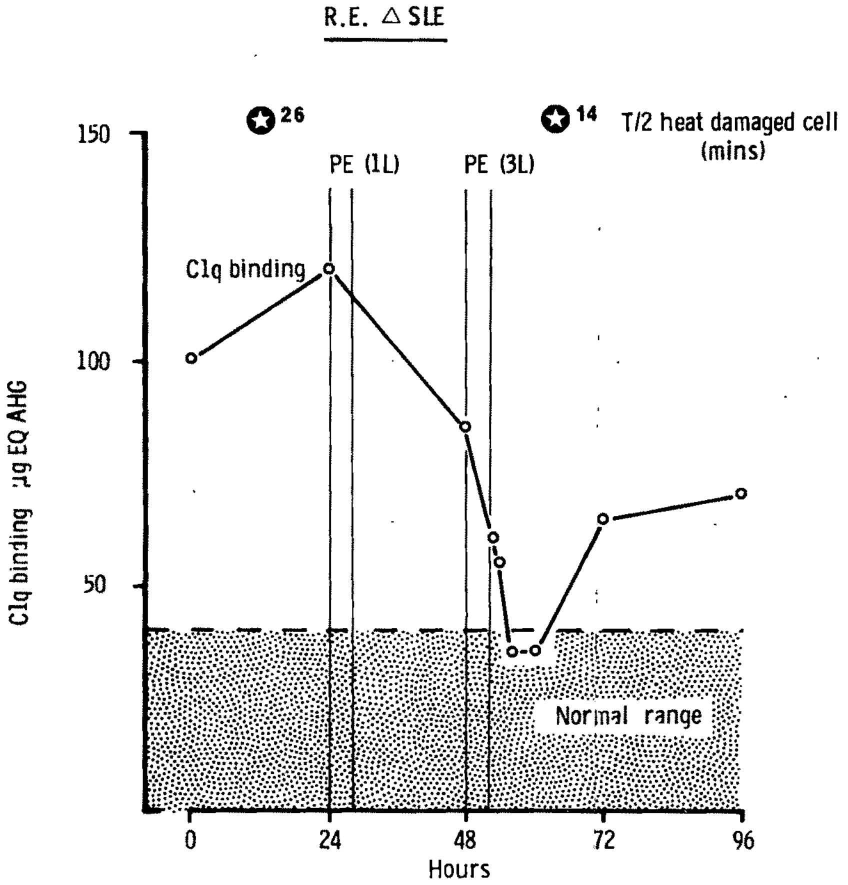


Figure 5 Serial measurements of Clq binding during two plasma exchanges in patient RE with SLE. (Normal serum binds less than 40 µg equivalent aggregated IgG.) Note complex levels continue to fall after exchange. Heat-damaged cell clearance is abnormal when detectable complexes are present.

erythematosus (SLE) Verrier-Jones et al (14) reported that plasma exchange seemed to be of benefit in those patients with fulminating disease and high level of circulating complexes; other researchers (23) have recorded similar results. In cutaneous vasculitis, plasmapheresis alone at intervals of 3–4 weeks has been successful in the management of certain patients (24). Persistent nephritis attributed to circulating complexes detectable in the circulation long after successful antibiotic treatment of infective endocarditis has also shown good response to plasma exchange (25). Studies of immune mechanisms in certain neoplastic conditions suggest that antibodies or immune complexes may be of pathogenetic significance. Preliminary reports suggest that plasma exchange may be of some help (26, 27). Further investigation is necessary before reaching firmer conclusions about the value of interfering with immune mechanisms whose pathogenetic contribution remains unclear.

Comment

As a direct method of manipulating the constitution of the vascular compartment (and presumably the cellular environment) plasmapheresis is the most effective and immediate form of treatment available. Its limitations are determined by the role which humoral factors play in the development of the disease, whether primarily (such as specific autoantibodies) or secondarily (such as the mediators of the inflammatory response, e.g. fibrinogen or complement). Clearly, as far as primary pathogenetic agents are concerned, rate and duration of synthesis, specificity, and affinity are all properties that affect contribution to control of the disease. Accurate monitoring techniques of the levels of such factors are therefore of great importance. Expense, supply of plasma substitute, and time involved restrict the use of intensive plasma exchange as a long-term therapeutic tool.

Turning to the effect of plasma exchange on impaired reticuloendothelial function, experimental evidence shows that if the RES is overloaded, deposition of factors such as immune complexes normally cleared by this system occurs at other sites (28). However, this situation is not easy to achieve: relatively large quantities of IC's are necessary to bring about even a short-lived blockade of the reticuloendothelial system. That this impairment can be demonstrated in a relatively large proportion of patients with nephritis or vasculitis is important and implies that in human disease, associated with presence of immune complexes, impaired reticuloendothelial function can lead to deposition of injurious complexes in other organs and perhaps to the manifestations of multisystem injury which are the hallmark of immune complex disease. The freeing of reticuloendothelial blockade by a short period of plasma exchange may allow recovery of phagocytic function such that remaining immune complexes are cleared normally.

Replacing the plasma removed with fresh plasma, rather than albumin solution, may consequently replace components deficient in the patient's native plasma. In this respect the claim that greater benefit may be accrued by use of fresh plasma in the plasmapheresis of patients with SLE is of interest (29). A similar explanation could account for the reported improvement in patients with thrombotic thrombocytopenic purpura undergoing plasma exchange (30). Other workers found comparable benefit from simple infusion of fresh-frozen plasma (31). Specific enrichment or depletion of naturally occurring inhibitors in replacement plasma is currently being explored in our laboratory.

Clearly a major difficulty in assessing the value of this expensive form of therapy is mounting appropriate control studies. In anti-GBM disease, we think the results together with the evidence of control of anti-GBM antibody production are sufficiently impressive to justify the approach. More difficulty exists in the IC disease group, because of its heterogeneity. However, in this situation, we are currently conducting a control trial, in the hope of establishing the value of this treatment.

ACKNOWLEDGMENTS

We thank Dr. C. B. Wilson (Scripps Clinic, La Jolla, California) for measurement of anti-GBM antibodies, and Dr. B. Pussell (Royal Postgraduate Medical School, London) for measurement of IC's. We are grateful to Ms. Sally Gibbons and Ms. Benita Roberts for performing plasma exchange.

Literature Cited

1. Dixon, F. J., Feldman, J. D., Vazquez, J. J. 1961. Experimental glomerulonephritis; the pathogenesis of a laboratory model resembling the spectrum of human glomerulonephritis. *J. Exp. Med.* 113:899–920
2. Weisberger, A. S., Moore, R. D., Schoenberg, M. D. 1966. Modification of experimental immune nephritis by chloramphenicol. *J. Lab. Clin. Med.* 67:58–69
3. Cochrane, C. G., Unanue, E. R., Dixon, F. J. 1965. A role of polymorphonuclear leucocytes and complement in nephrotoxic nephritis. *J. Exp. Med.* 122:99–116
4. Naish, P. F., Penn, G. B., Evans, D. J., Peters, D. K. 1972. The effect of defibrination on nephrotoxic serum nephritis in rabbits. *Clin. Sci.* 42:643–46
5. Naish, P. F., Thomson, N. M., Simpson, I. J., Peters, D. K. 1975. The role of polymorphonuclear leucocytes in the autologous phase of nephrotoxic nephritis. *Clin. Exp. Immunol.* 22:102–11
6. Koffler, D., Sandson, J., Carr, R., Kunkel, H. G. 1969. Immunologic studies concerning the lung in Goodpasture's syndrome. *Am. J. Pathol.* 54:292–306
7. Wilson, C. B., Marquardt, H., Dixon, F. J. 1974. Radioimmune assay (RIA) for circulating antiglomerular basement membrane antibodies. *Kidney Int.* 6:114A
8. Lerner, R. A., Glassock, R. J., Dixon, F. J. 1967. The role of antiglomerular basement membrane antibody in the pathogenesis of human glomerulonephritis. *J. Exp. Med.* 126:989–1004
9. Lockwood, C. M., Rees, A. J., Pussell, B., Peters, D. K. 1977. Experience of the use of plasma exchange in the management of potentially fulminating glomerulonephritis and S.L.E. *Exp. Hematol.* 5 (Suppl. 1):117–36
10. Lockwood, C. M., Amos, N., Peters D. K. 1979. Goodpasture's syndrome: radioimmunoassay for measurement of circulating anti-GBM antibodies. *Kidney Int.* 16:93(A)
11. Pussell, B. A., Lockwood, C. M., Scott, D. M., Pinching, A. J., Peters, D. K. 1978. Value of immune complex assays in diagnosis and management. *Lancet* 2:359–63
12. Ewan, P., Jones, H. A., Rhodes, C. G., Hughes, J. M. B. 1976. Detection of pulmonary haemorrhage with carbon monoxide uptake. Application in Goodpasture's syndrome. *N. Engl. J. Med.* 295:1391–96
13. Pinching, A. J., Travers, R., Hughes, G. R. V., Jones, T., Moss, S. 1978. Oxygen-15 brain scanning for detection of cerebral involvement in systemic lupus erythematosus. *Lancet* 1:898–900
14. Verrier-Jones, J., Cumming, R. J., Bucknall, R. C., Asplin, C. M., Fraser, I. D., Bothamley, J., Davis, P., Hamblin, T. 1976. Plasma exchange in the management of systemic lupus erythematosus? *Lancet* 1:709–11
15. Lockwood, C. M., Rees, A. J., Pinching, A. J., Pussell, B., Sweny, P., Uff, J., Peters, D. K. 1977. Plasma exchange and immunosuppression in the treatment of fulminating immune-complex crescentic nephritis. *Lancet* 1:63–67

16. Lockwood, C. M., Worrledge, O., Nicholas, A., Cotton, C., Peters, D. K. 1979. Reversal of impaired splenic function in patients with nephritis or vasculitis (or both) by plasma exchange. *N. Engl. J. Med.* 300:524–30
17. Lockwood, C. M., Rees, A. J., Pearson, T. A., Evans, E. J., Peters, D. K., Wilson, C. B. 1976. Immunosuppression and plasma-exchange in the treatment of Goodpasture's syndrome. *Lancet* 1:711–15
18. Pinching, A. J., Peters, D. K., Newsom Davis, J. 1976. Remission of Myasthenia Gravis following plasma-exchange. *Lancet* 2:1373–76
19. Ruocco, V., Rossi, A., Argenziano, G., Astarita, C., Alviggi, L., Fatzati, B., Papaleo, G. 1978. Pathogenicity of the intercellular antibodies of pemphigus and their periodic removal from the circulation by plasmapheresis. *Br. J. Dermatol.* 98:237–43
20. Cotterill, J. A., Barker, D. J., Millard, L. G., Robinson, E. A. 1978. Plasma exchange in the treatment of pemphigus vulgaris. *Br. J. Dermatol.* 98:243
21. Graham-Pole, J., Barr, W., Willoughby, M. L. N. 1977. Continuous flow plasmapheresis in the management of severe rhesus disease. *Br. Med. J.* 1185–88
22. Tursz, T., Preud'Homme, J.-L., Labaume, S., Matuchansky, C., Seligman, M. 1977. Autoantibodies to B lymphocytes in a patient with hypoimmunoglobulinemia. Characterization and pathogenic role. *J. Clin. Invest.* 60:405–11
23. Rossen, R. D., Hersh, E. M., Sharp, J. T., McCredie, K. B., Gyorkey, F., Suki, W. N., Ecknoyan, G., Reisberg, M. A. 1977. Effect of plasma exchange on circulating immune complexes and antibody formation in patients treated with cyclophosphamide and prednisone. *Am. J. Med.* 63:674–82
24. Lockwood, C. M., Pussell, B. A., Wilson, C. B., Peters, D. K. 1979. Plasma-exchange in nephritis. *Adv. Nephrol.* 8:383–418
25. Landwehr, D. M., Evans, P. S., Fisher, L. M., Ruddy, S. 1977. Removal of immune complexes by plasmapheresis in glomerulonephritis with subacute bacterial endocarditis. *Clin. Res.* 25:507A
26. Hersey, P., Isbister, J., Edwards, A., Murray, E., Adams, E., Biggs, J., Milton, G. W. 1976. Antibody-dependent cell-mediated cytotoxicity against melanoma cells induced by plasmapheresis. *Lancet* 1:825–28
27. Israel, L., Mannoni, P., Radot, E., Greenspan, E. 1976. Reactions immunitaires et tumorales a l'echange de plasma dans les cancers avances. *Nouv. Presse Med.* 5:433
28. Mannik, M., Arend, W. P. 1971. Fate of preformed immune complexes in rabbits and rhesus monkeys. *J. Exp. Med.* 134(S):19–31
29. Moran, C. J., Parry, H. F., Mowbray, J., Richards, J. D. M., Goldstone, A. H. 1977. Plasmapheresis in systemic lupus erythematosus. *Br. Med. J.* 1:1573–74
30. Bukowski, R. M., King, J. W., Hewlett, J. S. 1977. Plasmapheresis in the treatment of thrombotic thrombocytopenic purpura. *Blood* 50:413–17
31. Byrne, J. J., Khurana, M. 1977. Response of thrombotic thrombocytopenic purpura to fresh plasma infusion. *Clin. Res.* 25:335A

Ann. Rev. Med. 1980. 31:181–90
Copyright © 1980 by Annual Reviews Inc. All rights reserved

CLINICAL PHARMACOLOGY OF ACUTE ALLERGIC DISORDERS

❖7352

John C. Foreman, Ph.D., M.B., B.S.[1]
and Lawrence M. Lichtenstein, M.D.

Division of Clinical Immunology, Johns Hopkins University
School of Medicine at the Good Samaritan Hospital,
5601 Loch Raven Boulevard, Baltimore, Maryland 21239

INTRODUCTION

Acute allergic disorders include anaphylactic shock, asthma, allergic rhinitis, and a variety of skin conditions such as urticaria and eczema. The pathological mechanism of acute allergic diseases are well enough understood to formulate a rational approach to therapy. A brief description of acute allergic disorders is followed by a discussion of their clinical pharmacology and some new approaches to treatment. We aim to give insight into mechanisms of actions of clinically useful drugs and also to demonstrate where a knowledge of pathology could be utilized to develop new therapeutic approaches.

MECHANISM OF ALLERGIC DISORDERS

Antigenic Challenge

The initial event in the genesis of an allergic response is the access of the antigen to those sites in the immunological effector system that bear the specific receptors for antigen. This involves the passage of the antigen across mucosal surfaces, and little is known about the processes limiting or controlling such passage.

---

[1] Present address: Department of Pharmacology, University College London, London WC1E 6BT, United Kingdom.

0066-4219/80/0401-0181$01.00

Antibody Response

Receptors for antigen within the immunological system appear to be multiple and are probably present on B lymphocytes, T lymphocytes, and macrophages. Different receptors for a specific antigen probably exist on the various cell types that become involved in antibody generation and its control.

Interaction of antigen with B cells or their precursors initiates the differentiation of such cells into memory B cells and IgE-antibody-secreting plasma cells. In addition, antigen appears to interact with a population of T lymphocytes to generate T suppressor cells, which inhibit IgE production from triggered B cells (1). Another population of T cells become helper cells and augment the IgE–B cell response; helper T cells may be produced by interaction of the T cell precursors with antigen-stimulated macrophages (2). There is also evidence that the suppressor T cells may be antigen-specific or nonspecific (3, 4). To complement the cellular control of IgE production there appears to be a soluble factor capable of exerting suppressor effects. The exact nature of the factor is unknown but it is similar to β_2 microglobulin, and is not itself an antibody (5). It is important to realize that antigen generates IgG antibody as well as IgE antibody and the relative proportion of these classes may be influenced in several ways (6, 7), a fact that may be of some therapeutic significance, as discussed below.

The ultimate product of the cellular and humoral events that follow interaction of antigen with the immune system is antigen-specific IgE antibody. The IgE antibody secreted by the mature plasma cells appears in the circulation and also fixes to cells (8, 9). IgE is very selective in terms of its cellular binding site, binding only to tissue mast cells, circulating basophils, and some B lymphocytes.

Cellular Response

The specific receptor site on mast cells and basophils for IgE is similar and is of high affinity ($K_A = 10^{11}$ liters·mole^{-1}) (10). There are about 10^5 receptors on each cell (11) and the receptor has been extracted and purified from the cell membrane proteins (12, 13). Several experimental approaches support the hypothesis that cross-linking of the IgE receptors on the mast cell or basophil membrane signals those cells to release preformed mediators and to synthesize other mediators that are not stored preformed within the cell. For example, antigen must be at least bivalent to stimulate cells to release mediators (14, 15); anti-IgE will release histamine as long as both combining sites are joined (16), but the Fab' monomer anti-IgE will not activate mediator release. Chemically dimerized IgE will release histamine when it binds to the cell, but monomer will not (17). The most elegant evidence that receptor cross-linking is the cellular trigger comes from experiments with antireceptor antibody. It has been shown that $(Fab')_2$ antireceptor antibody releases histamine from mast cells without the participation of IgE whereas the monomer Fab' is not a sufficient trigger (18).

The cross-linking of IgE-receptors on the membrane is transduced into information useful to the cell for releasing and synthesizing mediators. The transduction appears to involve the opening of a calcium channel in the cell membrane, allowing calcium to enter the cell. A rise in free calcium concentration within the cell is thought to bring about fusion of granule and cell membranes to release preformed mediators by exocytosis of granules (19). Calcium entry may also initiate synthesis of those mediators that are not preformed (20).

The formation of calcium channels following receptor cross-linking involves biochemical changes in membrane phospholipids: phosphatidyl inositol turnover increases (21) and phospholipid methylation occurs (22). The control of the mediator release activated by calcium entry probably involves intracellular cyclic nucleotide levels, which in turn are influenced by the interaction of hormones such as adrenaline, adenosine, and histamine at specific receptors on the cell surface (23–25). Increased intracellular levels of cyclic AMP inhibit mediator release probably by inhibiting calcium gate formation (26). Analogous to other systems (27), cyclic AMP could inactivate calcium channels by stimulating a protein kinase that phosphorylates the channels.

Mediators

The classical mediator of allergic phenomena is histamine, which is stored, preformed, within mast cell and basophil granules. In addition to histamine, the spasmogenic "slow reacting substance of anaphylaxis" (SRS-A) is synthesized (28), and prostaglandins are generated (29). Both of these probably act as mediators and recent work suggests that SRS-A, like the prostaglandins, is a product of arachidonic acid metabolism (30). In addition to histamine, SRS-A and prostaglandins which stimulate smooth muscle and increase vascular permeability, the allergic reaction generates factors which cause chemotaxis of eosinophils (31) and neutrophils (32). Also, mast cells and basophils liberate a proteolytic enzyme that cleaves Hageman factor and leads to production of kinins (33).

The summary of our knowledge about the mechanism of allergic reactions given above and shown in Figure 1 forms the basis of the following review of current and novel approaches to the therapy of these reactions.

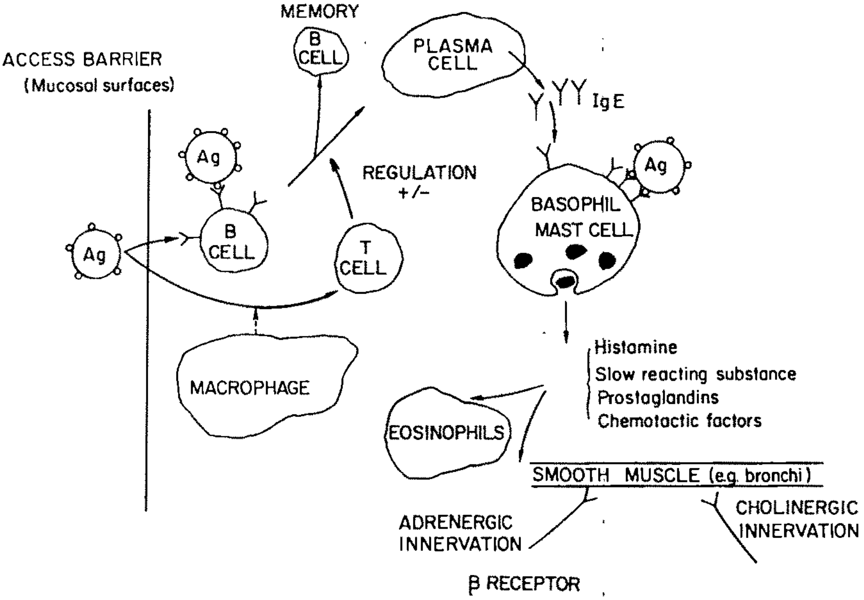


Figure 1 Diagrammatic representation of the major components of the sequence of events in the generation of an allergic response to antigen.

THERAPEUTIC APPROACHES IN ALLERGIC DISORDERS

Restricting possible contact of allergic individuals with the source of antigen forms an essential part of the management of allergic diseases, but at the present time there are no pharmacological tools for preventing the access of antigen to the immunological effector system.

Currently the most exciting area for the development of drugs for treating allergic diseases is the immunological effector system. From what has been said above, it will be clear that generation of suppressor T cells, inhibition of helper T cells, production of soluble suppressor factors, or the inactivation of IgE-producing cells are all theoretically feasible modes of interfering with allergic diseases. Some progress has been made in the generation of antigen-specific suppressor T cells by Takatsa & Ishizaka (34), who have shown in mice that the presentation of denatured antigen causes a decrease in serum IgE together with the generation of suppressor T cells that exert suppressor activity on IgE production. A clinical trial based on this approach involved individuals who were known to be allergic to ragweed pollen. Treatment of these subjects with urea-denatured antigen E of ragweed pollen led to a reduced IgE and IgG response compared with the response to nondenatured antigen (35). The lymphocytes of patients treated with denatured antigen showed impaired thymidine incorporation following antigen stimulation, but symptoms were not improved by the treatment.

Other experiments (36) in mice have shown that nonspecific ablation of T cells leads to an increase in IgE antibody responses but there is, at present, no method that could be tested in man for attempting to stimulate nonspecifically suppressor T cells. More promising perhaps is the observation that injection of adjuvants such as complete Freund's adjuvant into experimental animals leads to the generation of one or more soluble factors capable of suppressing the IgE antibody response (5). The exact nature of these factors is so far unknown, but they can be passively transferred from animal to animal and they are not immunoglobulins (37). There is some evidence that they may be associated with the plasma β_2 microglobulins. Again, the human equivalent of these suppressor factors has not been explored but it is attractive to speculate that knowledge of the chemistry of such factors could form the basis of a drug development program for the suppression of IgE production in response to allergenic stimulation.

Attempts to inactivate the B cells or plasma cells that lead to IgE secretion have proved successful in experimental animals and we await the clinical application of this approach. B cells are inactivated through use of haptenic groups chemically conjugated with a nonimmunogenic polymer of D-glutamic acid and D-lysine (38). Cells previously sensitized for IgE antibody production are rendered specifically and irreversibly insensitive by treatment with these haptenic conjugates. The method appears to involve only B cell inactivation and no modulation of T cell response (38).

Once antibody has been produced, the antigen-antibody reaction that triggers the cells to release mediators can be prevented in two ways. One way is to block the binding of the antibody to the cells; in this regard a recent report that a peptide fragment of IgE could inhibit allergic skin response to antigen is of interest (39). It is conceivable that a polypeptide of similar structure to the Fc region of the IgE molecule could occupy Fc receptors on mast cells and prevent the binding of antigen specific IgE. The original observation was not confirmed (40) and again this remains an approach for further work.

A second means of blocking the antigen-antibody reaction is immunotherapy, which is thought to prevent antigen binding to cell-fixed IgE. But immunotherapy is of proven value only in some allergic diseases, for example allergic rhinitis due to pollen antigens, but in other instances such as asthma this therapy is not clinically efficacious. Blocking antibodies of the IgG class are supposed to bind the antigen and prevent its access to mast cells. The blocking antibody hypothesis is not universally accepted as the explanation of the clinical efficacy of immunotherapy but it remains an attractive model.

Using the recent report of the clinical value of immunotherapy with a specific protein molecule, investigators can now rigorously test the blocking antibody hypothesis. Lichtenstein and colleagues (41) showed that injections of phospholipase from bee venom leads to IgE and IgG antibody responses, and patients sensitive to bee stings are entirely protected from anaphylaxis by this treatment. Marsh and colleagues (42) used formaldehyde-treated allergens called allergoids (cf toxoids) to generate excess IgG relative to IgE antibodies, thereby stimulating blocking antibody production and reducing sensitivity to allergen. The trial of this approach in man continues but the initial results suggest that while it is effective, the use of allergoid to "desensitize" patients is only marginally better than allergen.

The pharmaceutical industry has actively attempted to develop drugs that inhibit the release of mediators by acting directly on mast cells or basophils ever since cromoglycate was discovered (43). The mediator release described suggests that drugs which elevate cellular cyclic AMP (β-adrenergic receptor agonists, phosphodiesterase inhibitors) should suppress mediator release. Another method of reducing release of mediators would be to prevent calcium transport into mast cells: indeed, elevation of cyclic AMP levels may do just that (44). Activation of a β-receptor in human lung is certainly capable of inhibiting mediator release by antigen-IgE stimulation (45), and this may account in part for the clinical effectiveness of β-agonists in the therapy of asthma. It is presumed that β-receptor stimulation activates adenyl cyclase and causes a rise in intracellular levels of cyclic AMP.

Theophylline is similarly active in preventing mediator release and may do so by virtue of inhibiting the phosphodiesterase enzyme responsible for destroying cyclic AMP. However, theophylline is an antagonist of adenosine at those concentrations at which it exerts its clinical effect (46); this mechanism might account in part for its suppression of allergic diseases, since adenosine increases histamine release from mast cells (47). It should be noted, however, that in human basophils adenosine itself inhibits allergic mediator release (25).

Cromoglycate belongs to a novel class of drugs that also appear to act by inhibiting mediator release. (48). It has the apparent disadvantage of only being active when administered by inhalation. It is noteworthy that many thousands of compounds screened for cromoglycate-like activity in rat mast cells, and subsequently found to be antiallergic when administered orally to experimental animals, have failed to yield a new drug for clinical use. The mode of action of cromoglycate remains unknown but there is evidence that it blocks the antigen-stimulated transport of calcium across the mast cell membrane (49). The drug is also a phosphodiesterase inhibitor but the low potency of this action suggests that it is unrelated to the clinical effect of the drug (50).

The final step in the generation of an allergic response is the action of mediators on target organs. Mediator antagonists currently form the most powerful battery of drugs for the clinical management of allergic diseases. Evidence suggests that antagonists to several mediators must be employed to abolish an allergic response: histamine, SRS-A, and prostaglandins apparently contribute the most. Antihistamines acting on H_1 receptors can abolish the allergic response in some animal species but in man they are, at best, only partly effective in suppressing allergic rhinitis. In asthma both H_1 and H_2 antihistamines are ineffective (51). In experimental models the combination of SRS-A antagonist and antihistamine blocks most of the bronchial smooth muscle response to an antigen-antibody reaction (52). The development of the SRS-A antagonist FPL 55712 is clearly a major advance (53), and the recent description of the structure of SRS-A (54) should lead to more rapid progress toward clinically useful SRS-A antagonists.

Nonspecific antagonists of the actions of mediators, for example, β_2-receptor agonists and phosphodiesterase inhibitors, are the mainstay of allergic disease therapy. Steroids, which are clinically the most potent drugs used to treat allergic diseases, also appear to act predominantly by suppressing the effects of released mediators, although their precise mechanism of action is unknown. It is interesting that steroids are of little or no value in anaphylaxis but are very potent in chronic allergic diseases. A significant advance in steroid therapy has been the production of very potent steroids such as beclamethasone in a form for topical administration by inhalation or spray (55). Such topical administration has marked clinical efficacy while reducing to a very large degree the likelihood of side effects from steroid administration (55).

It is interesting that transfer from oral to inhaled therapy with steroids has yielded such a drastic advance for clinical medicine while the drug houses have so far been unable to develop an orally active cromoglycate-like compound. Allergic diseases are usually manifest locally, and so locally directed pharmacology may well prove to be of greatest benefit to patients.

ACKNOWLEDGMENTS

J.C.F. is in receipt of a Fulbright-Hays Travel Award. Support for this work comes from grant AI 07290 from the Institute of Allergy and Infectious Diseases and grant HL 23586 of the Heart and Lung Institute of the National Institutes of Health.

Literature Cited

1. Dixon, F. J., McConahey, R. W. 1963. Enhancement of antibody formation by whole body X-radiation. *J. Exp. Med.* 117:833–45
2. Ishizaka, K., Adachi, T. 1976. Generation of specific helper cells and suppressor cells *in vitro* for the IgE and IgG antibody responses. *J. Immunol.* 117:40–47
3. Chiorazzi, N., Fox, D. A., Katz, D. M. 1977. Hapten-specific IgE antibody responses in mice. VII. Conversion of IgE "non-responder" strains to IgE "responders" by elimination of suppressor T cell activity. *J. Immunol.* 118:48–54
4. Watanabe, N., Kojima, S., Shen, F. W., Ovary, Z. 1977. Suppression of IgE antibody production in SJL mice. II. Expression of Ly-1 antigen on helper and non-specific suppressor T cells. *J. Immunol.* 118:485–88
5. Tung, A. S., Chiorazzi, N., Katz, D. M. 1979. Regulation of IgE antibody production by serum molecules. I. Serum from complete Freund's adjuvant–immune donors suppresses irradiation-enhanced IgE production in low responder mouse strains. *J. Immunol.* 120:2050–59
6. Mota, I. 1965. The mechanism of anaphylaxis. I. Production and biological properties of "mast cell sensitizing antibodies." *Immunology* 7:681–99
7. Katz, D. M. 1978. Control of IgE antibody production by suppressor substances. *J. Allergy Clin. Immunol.* 62:44–55
8. Tomioka, M., Ishizaka, K. 1971. Mechanisms of passive sensitization. II. Presence of receptors for IgE on monkey mast cells. *J. Immunol.* 107:971–78
9. Ishizaka, K., Tomioka, M., Ishizaka, T. 1970. Mechanisms of passive sensitization. I. Presence of IgE and IgG molecules on human leukocytes. *J. Immunol.* 105:1459–67
10. Kulczycki, A., Metzger, H. 1974. The interaction of IgE with rat basophilic leukemia cells. II. Quantitative aspects of the binding reaction. *J. Exp. Med.* 140:1670–95
11. Ishizaka, T., König, W., Kurata, M. Mauser, L., Ishizaka, K. 1975. Immunologic properties of mast cells from rats infected with *Nippostrongylus brasiliensis. J. Immunol.* 115:1078–83
12. Conrad, D. H., Froese, A. 1976. Characterization of the target cell receptor for IgE. II. Polyacrylamide gel analysis of the surface IgE receptor from normal rat mast cells and from rat basophilic leukemic cells. *J. Immunol.* 116:319–26
13. Carson, D. A., Kulczycki, A., Metzger, H. 1975. Interaction of IgE with rat basophilic leukemia cells. III. Release of intact receptors on cell free particles. *J. Immunol.* 114:155–60
14. Levine, B. B. 1965. The nature of the antigen-antibody complexes which initiate anaphylactic reactions. I. Quantitative comparison of the abilities of non-toxic univalent, toxic univalent, divalent and multivalent benzyl penicilloyl haptens to evoke passive cutaneous anaphylaxis in the guinea pig. *J. Immunol.* 94:111–20
15. Magro, A., Alexander, A. 1974. *In vitro* studies of histamine release from rabbit leukocytes by divalent haptens. *J. Immunol.* 112:1757–61
16. Ishizaka, T., Ishizaka, K., Johansson, S. G. O., Bennich, H. 1969. Histamine release from human leukocytes by anti IgE antibodies. *J. Immunol.* 102:884–92
17. Segal, D. M., Taurog, J. D., Metzger, H. 1977. Dimeric immunoglobulin E serves as a unit signal for mast cell degranulation. *Proc. Natl. Acad. Sci. USA* 74:2993–97
18. Ishizaka, T., Ishizaka, K. 1978. Triggering of histamine release from rat mast cells by divalent antibodies against IgE-receptors. *J. Immunol.* 120:800–5
19. Foreman, J. C., Garland, L. G., Mongar, J. L. 1976. The role of calcium in secretory processes: model studies in mast cells. In *Calcium Transport in Contraction and Secretion,* ed. E. Carafoli, F. Clementi, W. Drabikowski, A. Margreth, pp. 175–84. Amsterdam: North-Holland
20. Bach, M. K., Brashler, J. D. 1976. *In vivo* and *in vitro* production of a slow reacting substance in the rat upon treatment with calcium ionophore. *J. Immunol.* 113:2040–44
21. Cockcroft, S., Gomperts, B. D. 1979. Evidence for a role of phosphatidyl inositol turnover in stimulus-secretion coupling. *Biochem. J.* 188:681–87
22. Hirata, F., Axelrod, J., Crews, R. T. 1979. Concanavalin A stimulates phospholipid methylation and phosphatidylserine decarboxylase in rat mast cells. *Proc. Natl. Acad. Sci. USA.* In press
23. Assem, E. S. K., Schild, H. O. 1971. Antagonism by β-adrenergic blocking agents of the antianaphylactic effect of isoprenaline. *Br. J. Pharmacol.* 62:620–30

24. Lichtenstein, L. M., Gillespie, E. 1975. The effects of H1 and H2 antihistamines on "allergic" histamine release and its inhibition by histamine. *J. Pharmacol. Exp. Ther.* 192:441–50
25. Marone, G., Findlay, S. F., Lichtenstein, L. M. 1979. Adenosine receptor on human basophils: modulation of histamine release. *J. Immunol.* In press
26. Foreman, J. C., Hallett, M. B., Mongar, J. L. 1977. The relationship between histamine secretion and 45-calcium uptake in rat mast cells. *J. Physiol. London* 271:193–214
27. Langan, T. A. 1972. Protein kinases and protein kinase substrates. *Adv. Cyclic Nucleotide Res.* 3:99–153
28. Brocklehurst, W. E. 1960. The release of histamine and formation of a slow reacting substance (SRS-A) during anaphylactic shock. *J. Physiol. London* 151:416–35
29. Piper, P. J., Vane, J. R. 1969. Release of additional factors in anaphylaxis and its antagonism by anti-inflammatory drugs. *Nature* 223:29–35
30. Jakschik, B. A., Falkenhein, S., Parker, C. W. 1977. Precursor role of arachidonic acid in release of slow reacting substance from rat basophilic leukemia cells. *Proc. Natl. Acad. Sci. USA* 74: 4577–81
31. Czarnetzki, B. M., König, W., Lichtenstein, L. M. 1978. Antigen-induced eosinophil chemotactic factor (ECF) release by human leukocytes. *Inflammation* 1:201–5
32. Ward, P. A., Newman, L. J. 1969. A neutrophil chemotactic factor from human C'5. *J. Immunol.* 102:93–99
33. Newball, H. H., Talamo, R. C., Lichtenstein, L. M. 1975. Release of leukocyte kallikrein mediated by IgE. *Nature* 254:635–36
34. Takatsa, K., Ishizaka, K. 1976. Reaginic antibody formation in the mouse. VII. Induction of suppressor T cells for IgE and IgG antibody responses. *J. Immunol.* 116:1257–64
35. Norman, P. S., Marsh, D. G., Ishizaka, K., Lichtenstein, L. M. 1976. New immunologic methods of treatment in IgE mediated allergies. In *Allergy and Clinical Immunology. Excerpta Medica Int. Cong. Ser.* 414, pp. 482–86
36. Chiorazzi, N., Fox, D. A., Katz, D. M. 1976. Hapten-specific IgE antibody responses in mice. VI. Selective enhancement of IgE antibody production by low doses of X-irradiation and by cyclophosphamide. *J. Immunol.* 117: 1629–37
37. Katz, D. M., Tung, A. S. 1978. Regulation of IgE antibody production by serum molecules. II. Strain specificity of the suppressive activity of serum from complete Freund's adjuvant–immune low responder mouse donors. *J. Immunol.* 120:2060–67
38. Katz, D. M., Davie, J. M., Paul, W. E., Benacerraf, B. 1971. Carrier function in anti-hapten antibody responses. IV. Experimental conditions for the induction of hapten specific tolerance and for the stimulation of anti-hapten anamnestic responses by non-immunogenic hapten-polypeptide conjugates. *J. Exp. Med.* 134:201–23
39. Hamburger, R. N. 1975. Peptide inhibition of the Prausnitz-Kustner reaction. *Science* 189:389–90
40. Bennich, M., Ragnarsson, U., Johansson, S. G. O., Ishizaka, K., Ishizaka, T., Levy, D. A., Lichtenstein, L. M. 1977. Failure of the putative IgE pentapeptide to compete with IgE for receptors on basophils and mast cells. *Int. Arch. Allergy Appl. Immunol.* 58:459–64
41. Hunt, K. J., Valentine, M. D., Sobotka, A. K., Benton, A. W., Amodio, F. J., Lichtenstein, L. M. 1978. A controlled trial of immunotherapy in insect hypersensitivity. *N. Engl. J. Med.* 299:157–61
42. Marsh, D. G., Lichtenstein, L. M., Campbell, U. D. H. 1970. Studies on "allergoids" prepared from naturally-occurring allergens. I. Assay of allergenicity and antigenicity of formalized rye Group 1 allergen. *Immunology* 18:705–20
43. Garland, L. G., Green, A. F., Hodson, H. 1978. Inhibitors of the release of anaphylactic mediators. In *Inflammation,* ed. J. R. Vane, S. H. Ferreira. *Handbook of Experimental Pharmacology* 50(2), Chap. 34. Berlin, Heidelberg, New York: Springer
44. Foreman, J. C., Mongar, J. L., Gomperts, B. D., Garland, L. G. 1975. A possible role for cyclic AMP in the regulation of histamine secretion and the action of cromoglycate. *Biochem. Pharmacol.* 24:538–40
45. Assem, E. S. K., Schild, H. O. 1969. Inhibition by sympathomimetic amines of histamine release by antigen in passively sensitized human lung. *Nature* 224:1028–29
46. Fredholm, B. B., Sydbom, A. 1980. Are the antiallergic actions of theophylline due to antagonism at the adenosine receptor? *Agents and Actions.* In press

47. Marquardt, D. L., Parker, C. W., Sullivan, T. J. 1978. Potentiation of mast cell mediator release by adenosine. *J. Immunol.* 120:871–78
48. Cox, J. S. G., Beach, J. E., Blair, A. M. J. N., Clarke, A. J., King, Lee, T. C., Loveday, D. E. E., Moss, G. F., Orr, T. S. C., Ritchie, J. T., Sheard, P. 1970. Disodium cromoglycate. *Adv. Drug Res.* 5:115–96
49. Foreman, J. C., Hallett, M. B., Mongar, J. L. 1977. Site of action of the antiallergic drugs cromoglycate and doxantrazole. *Br. J. Pharmacol.* 59:473–74P
50. Roy, A. C., Warren, B. T. 1975. Inhibition of cAMP phosphodiesterase by disodium cromoglycate. *Biochem. Pharmacol.* 23:917–19
51. Eiser, N. 1980. In *The Mast Cell,* ed. J. Pepys. London: Pitman Medical. In press
52. Adams, G. K., Lichtenstein, L. M. 1979. *In vitro* studies of antigen-induced bronchospasm: effects of antihistamines and SRS-A antagonist on response of sensitized guinea pig and human airways to antigen. *J. Immunol.* 122:555–82
53. Augstein, J., Farmer, J. B., Lee, T. B. Sheard, P., Tattersall, M. L. 1973. Selective inhibition of slow reacting substance of anaphylaxis. *Nature* (*New Biol.*) 245:215–17
54. Samuelsson, B. I. 1980. Oxygenation and further transformation of polyunsaturated fatty acids. Proc. 4th Int. Prostaglandins Cong. In *Adv. Prostaglandin Thromboxane Res. Ser.,* Vol. 6 and 7, New York: Raven Press. In press
55. Brown, H. M., Storey, G., George, W. H. S. 1972. Beclamethasone diproprionate: a new steroid aerosol for the treatment of allergic asthma. *Br. Med. J.* 1:585–90

Ann. Rev. Med. 1980. 31:191–200
Copyright © 1980 by Annual Reviews Inc. All rights reserved

AUTOIMMUNE GRANULOCYTOPENIA

❖7353

G. L. Logue, M.D., and D. S. Shimm, M.D.

Section of Hematology and Medical Oncology, Department of Medicine,
Durham Veterans Administration and Duke University Medical Centers,
Durham, North Carolina 27705

INTRODUCTION

Human blood cells are the targets of autoimmune reactions in a variety of diseases. In vitro immunologic testing is useful in clinically defining immune hemolytic anemia and, more recently, immune thrombocytopenia. Such testing is useful not only to elucidate pathophysiology, but to plan appropriate therapy as well. In contrast, the understanding of autoimmune reactions involving human granulocytes is somewhat more primitive. Although in vitro antigranulocyte antibody assays were described more than 20 years ago, techniques for the study of immune reactions involving granulocytes have been slow to develop and somewhat difficult to standardize.

However, compelling evidence suggests that immune reactions directed against granulocytes cause granulocytopenia in some patients. Placental transfer of isoimmune antibodies produces granulocytopenia in patients with neonatal neutropenia, and drug-related antibodies produce granulocytopenia in patients with some forms of drug-induced agranulocytosis (1, 2). Furthermore, the association of granulocytopenia with collagen vascular diseases suggests that autoimmune granulocytopenia, in addition to autoimmune hemolytic anemia and immune thrombocytopenia, occurs in this patient population (3). A variety of in vitro procedures can detect antibodies directed against granulocytes in these patients. We first describe the types of in vitro assays used to study autoimmune neutropenia, then relate these laboratory findings to the clinical syndromes where autoimmune granulo-

0066-4219/80/0401-0191$01.00

cytopenia appears. Finally, we speculate upon the potential usefulness of such testing in the therapy of autoimmune granulocytopenia.

GRANULOCYTE ANTIBODY TESTING

A variety of granulocyte antibody assays have been used to study patients with suspected autoimmune neutropenia. Whether the assay is immunochemical or functional, and whether the assay measures antibody on the patients' own granulocytes (direct) or in patients' serum (indirect) distinguishes these procedures.

Immunochemical Assays

These assays detect immunoproteins on the surface of granulocytes, either directly or indirectly. Leukoagglutinin techniques, indirect assays that will detect granulocyte alloantibodies, are useful in identification of drug-induced immune neutropenia and neonatal neutropenia (1, 2). These procedures are, however, often difficult to standardize (4). Lalezari and co-workers (5) described an infant with chronic neutropenia in whom autoantibodies directed against the neutrophil antigen NA2 were detected by leukoagglutination and reacted with her own granulocytes. Other studies of granulocyte aggregation suggest that leukocyte-leukocyte interaction may be stimulated by complement activation (6).

Antigranulocyte antibodies have been detected by a variety of procedures that utilize anti-immunoglobulin antisera. We developed assays for quantitating the amount of IgG bound to a patient's own granulocytes and the amount of the serum IgG that will bind to normal granulocytes using an antiglobulin consumption procedure (7, 8). In this assay, antihuman IgG is incubated with leukocytes or known concentrations of fluid phase IgG; following incubation, the amount of anti-IgG remaining is detected by complement lysis of sheep erythrocytes coated with human IgG. This assay has proven most useful in the study of granulocytopenia associated with rheumatoid arthritis, the results of which are described in detail below.

Rosenthal et al (9) in 1974 also used an antiglobulin consumption assay to test for antileukocyte IgG in patients with Felty's syndrome. They found these antibodies in 13 of 15 such patients, and concluded that antileukocyte IgG was specifically directed against granulocytes.

Fab anti-F(ab')$_2$ consumption assays have also been used to detect granulocyte autoantibodies (3). Starkebaum and colleagues described a patient with systemic lupus erythematosus (SLE) and neutropenia in whom such antibodies were detected. Using both in vitro and in vivo measurements, these authors related this antibody to granulocyte destruction. They showed granulocyte binding of intermediate and large IgG-containing molecular complexes as well as monomeric IgG. All of the immunochemical assays described here will detect granulocyte binding of IgG-containing immune complexes as well as binding of monomeric IgG antigranulocyte antibodies.

Calabresi et al (10) and, more recently, Verheugt et al (11) used fluorescent antiglobulin antisera to detect antigranulocyte antibodies. The latter authors used a continuous flow microfluorimeter to quantitate antiglobulin binding. With these technologic advances, they could accurately measure both direct and indirect granulocyte-bound immunoglobulins.

Functional Granulocyte Antibody Assays

The above mentioned granulocyte antibody assays are immunochemical since they detect immunoglobulins bound to leukocytes. Another class of procedures tests for in vitro pathophysiologic sequelae of granulocyte antibodies. Boxer & Stossel (12) studied the effects of antihuman neutrophil antibodies in vitro. Low concentrations of IgG stimulated ingestion of particles by neutrophils, which higher concentrations inhibited. These authors also developed an indirect serologic test for granulocyte antibodies utilizing rabbit macrophages to detect sensitized human leukocytes (4). In this assay, the rate of reduction of the dye, nitroblue tetrazolium, by rabbit macrophages was used to quantitate granulocyte antibody binding. In a similar fashion, antineutrophil antibodies can be quantitated by an opsonization assay in which activation of nonsensitized human neutrophils occurs when co-cultured with antibody-coated leukocytes (2). This assay has been used to study drug-induced immune neutropenia.

More recently, granulocyte antibodies have been detected by the binding of staphylococcal protein A (13). This bacterial product binds to Fc fragments of immunoglobulins of subclasses 1, 2, and 4 and thus can be used to detect cell-bound antibody. This reagent would not detect monomeric IgG bound to granulocytes by the granulocyte Fc receptor but would detect cell-bound aggregated IgG as well as antibodies directed against granulocyte antigens (14).

Rabbit complement-dependent microcytotoxicity assays have been used to detect human granulocyte autoantibodies (15, 16). This type of granulocyte antibody assay is most useful for detection of cold reacting autocytotoxins of IgM subclass. Interestingly, this granulocyte antibody assay not only detects autoantibodies in about one half of SLE patients, but positive reactions are also found in sera of approximately 10% of nonimmunized normals (15).

In this laboratory, we have developed a functional assay for granulocyte antibodies in which antibody-dependent lymphocyte-mediated granulocytotoxicity (ADLG) is detected (17, 18). These IgG antibodies bind to granulocytes and enable certain lymphocytes to lyse the sensitized cells. This reaction depends on monomeric antigranulocyte IgG. The clinical significance of such antibodies is described below.

ASSOCIATED DISEASES

Rheumatoid Arthritis

Felty (19) described the syndrome of rheumatoid arthritis, splenomegaly, and granulocytopenia in 1924. Since then, indirect evidence of an underlying autoimmune mechanism in Felty's syndrome has accumulated, including its occasional association with autoimmune hemolytic anemia, its response to corticosteroids or splenectomy, and the production of neutropenia by passive transfer of Felty's serum (10, 20–22). Early attempts to demonstrate directly immune neutropenia in Felty's syndrome were hindered by methodologic difficulties.

We studied 35 patients with seropositive rheumatoid arthritis and granulocytopenia (<2500 per mm^3) for granulocyte-bound IgG using the antiglobulin consumption assay described above (Table 1) (7, 8). The patients were classified by whether their granulocytopenia was mild (1000–2500 per mm^3) or severe (<1000 per mm^3), whether their spleen was enlarged, and whether they had increased granulocyte-bound IgG. Seventeen of these patients had severe granulocytopenia, splenomegaly, and elevated granulocyte-bound IgG. Eleven patients had moderate granulocytopenia without splenic enlargement and had normal levels of granulocyte-bound IgG. Some of these patients appeared to have drug-related granulocytopenia (gold, phenylbutazone, cytotoxic agents). This group of patients apparently had moderate granulocytopenia not related to pathologic IgG. Other possible explanations for granulocytopenia in these patients include a shift of granulocytes from "circulating" to "marginal" pools or

Table 1 Clinical parameters and granulocyte-bound IgG in patients with rheumatoid arthritis and granulocytopenia[a]

Granulocyte count < 1000 per mm^3	Splenomegaly	Elevated granulocyte-bound IgG	Number of patients
Yes	Yes	Yes	17
Yes	No	Yes	2
No	Yes	Yes	1
Yes	Yes	No	2
No	Yes	No	2
No	No	No	11

[a] Absolute granulocyte count less than 2500 per mm^3.

defective production of granulocytes. Granulocyte kinetic studies done in patients with rheumatoid arthritis have shown that both such mechanisms may be active (23, 24). Decreased granulocyte production does not exclude an immune basis for the granulocytopenia, since granulocyte autoantibodies could also affect maturing bone marrow precursors, analogous to the autoantibodies directed against erythrocyte precursors in patients with pure red cell aplasia. There is evidence that bone marrow granulocyte production may be suppressed in Felty's syndrome patients by cell-mediated mechanisms as well as by antigranulocyte antibodies (25, 26).

The small groups of patients outlined in Table 1 require explanation. The two patients in Row 4 represent false negative results for the granulocyte-bound IgG assay since they had severe granulocytopenia and splenomegaly without abnormal granulocyte-bound IgG. Whether this represents granulocytopenia caused by non-IgG antibodies or by nonimmune mechanisms remains to be proven. The two patients in Row 2 had rheumatoid arthritis, severe granulocytopenia, and elevated granulocyte-bound IgG without splenomegaly. Studies of granulocyte antibodies in these patients suggest that they have "Felty's syndrome without splenomegaly"(8). One of these patients, responded well to splenectomy. Thus, this type of immunochemical assay is useful in defining specific disease processes.

Systemic Lupus Erythematosus

Along with thrombocytopenia and autoimmune hemolytic anemia, leukopenia is noted in approximately 50% of patients with SLE (27). Early studies, using antiglobulin consumption assays, demonstrated leukocyte-bound antibody in most SLE patients without any correlation to clinical leukopenia (28). More recently, microcytotoxicity studies have shown nearly universal occurrence of cold reactive IgM antilymphocyte antibodies (29). The pathogenesis of granulocytopenia in SLE is unclear. Multiple mechanisms including decreased marrow production, increased peripheral destruction, and abnormal partitioning of granulocytes may be responsible (23).

Verheugt et al (11) described a leukopenic patient with SLE whose neutrophils carried IgG as measured by the direct immunofluorescent assay and whose serum contained neutrophil-specific IgG detected similarly. Following treatment with prednisone and azathioprine, her granulocyte count normalized and the neutrophil antibody disappeared. Starkebaum et al (3) described a neutropenic SLE patient whose serum contained three times the normal amount of granulocyte-binding IgG as measured by their antiglobulin consumption assay and whose serum had twice normal granulocyte-opsonizing ability. Neutrophil kinetics revealed a shortened intravascular neutrophil survival and an increased neutrophil mitotic pool, which, with the failure to respond to a neutropenic stimulus with increased neutrophil delivery, implied ineffective granulopoiesis. Serum fractionation localized IgG neutrophil-binding activity to fractions containing immune complexes and to monomeric IgG, while opsonizing activity was confined to monomeric IgG alone. Following treatment with prednisone and cyclophosphamide, granulocyte kinetics returned to normal while serum granulocyte-binding IgG and opsonizing activity remained abnormally elevated.

Several conclusions can be drawn from these observations. First, while the neutropenia of SLE is clearly multifactorial, immune mechanisms contribute in at least a fraction of these patients. Second, immune neutropenia in SLE might be found more commonly if antibodies directed against granulocyte precursors were also sought. Third, antibodies, found in patients with SLE, appeared to be pathogenetically significant rather than an epiphenomenon, as neutropenia disappeared with resolution of at least one immunologic abnormality in each case described. Understanding of the pathogenetic role of these antibodies will be valuable for planning therapy as described below.

Other Clinical Associations

Granulocyte autoantibodies are seen in a variety of patients without collagen vascular diseases. In some patients, this type of autoantibody is "idiopathic" and the patient with autoimmune neutropenia due to anti-NA2 antibodies, described by Lalezari (5), did not have other identifiable diseases. Likewise, numerous patients with idiopathic neutropenia have been found to have antigranulocyte antibodies detected by neutrophil opsonization and by immunofluorescent antiglobulin binding (4; 11). We also studied a variety of patients with granulocytopenia and elevated granulocyte-bound IgG who have no other obvious associated disease (30).

Weitzman & Stossel (2) described persistent autoimmune granulocytopenia precipitated by drug exposure. Following the ingestion of quinidine, chlorothiazide, and chlorpromazine, the patient developed granulocytopenia and had antineutrophil IgG. The antibodies persisted after the drugs were discontinued and appeared to react directly with granulocytes. These authors draw the analogy to methyldopa-related hemolytic anemia.

As with autoimmune hemolytic anemia and immune thrombocytopenia, autoimmune granulocytopenia appears to be associated with lymphoproliferative disorders. Verheugt et al (11), using an immunofluorescent granulocyte antibody assay, found four patients with angioimmunoblastic lymphadenopathy who had granulocyte-bound antibodies. Three of their patients had both IgG and IgM while the fourth had only IgA autoantibodies. Antiglobulin consumption assays, such as that used in this laboratory, detect only IgG, and thus would not detect this latter autoimmune reaction.

We studied a patient with chronic lymphocytic leukemia who had an antigranulocyte antibody that supports ADLG (30). This CLL patient was suspected of having antigranulocyte antibodies because of profound granulocytopenia with granulocyte precursors in the bone marrow. Four other CLL patients with granulocytopenia associated with extensive marrow replacement by leukemic lymphocytes did not show this granulocyte antibody activity.

We also found that serum from two granulocytopenic patients with a history of immune thrombocytopenic purpura (ITP) supports ADLG (30). Both these patients had previously been treated by splenectomy for chronic ITP. Following a remission of ITP, they developed granulocytopenia and this antigranulocyte antibody was detected in their serum. Other authors have also noted patients with granulocyte autoantibodies in association with platelet and red cell autoantibodies (11). It should be pointed out here that Evans & Duane (31) predicted in 1949 that this sort of autoimmune reaction would occur.

TREATMENT

As might be expected in a syndrome with diverse causes, uncertain pathophysiology, and infrequent incidence, the proper treatment of immune neutropenia is unclear. Additionally, indications for treatment of autoimmune neutropenia are not well defined. Unlike immune thrombocytopenia or hemolytic anemia where immune cytopenia dictates specific interventions such as splenectomy or corticosteroid therapy, severe degrees of immune neutropenia may be tolerated without serious infection (32). Although controversy exists, it would seem prudent to require severe or recurrent infections as well as granulocytopenia before undertaking specific therapy for autoimmune neutropenia.

Splenectomy

Splenectomy is the mainstay of therapy for autoimmune neutropenia, with up to 60% of patients with Felty's syndrome sustaining increases in granulocyte counts to greater than 4000 cells per mm^3 and decreased infections in one series (22). Following splenectomy, the amount of granulocyte-bound IgG in Felty's syndrome patients is nearly normal, with lower but still abnormal granulocyte-binding IgG in the serum (7, 8). On the other hand, the hazards of splenectomy must be considered, including increased susceptibility to infection, which caused death in one patient described in a series of Felty's patients (33) and increased incidence of infections and vasculitic complications in a series of SLE patients (34). Moreover, in Felty's patients with recurrent granulocytopenia post-splenectomy, we have demonstrated ADLG (18). In one patient, this antibody activity developed as granulocytopenia recurred nine weeks after splenectomy. Thus, development of this type of antibody activity appears to be a mechanism of splenectomy failure in some patients.

Corticosteroid Therapy

Corticosteroid therapy has been responsible for major improvement in some patients with autoimmune granulocytopenia (3, 5, 11). This success must be tempered by the fact that corticosteroids can cause shifts in granulocyte pools with apparent increases in granulocyte counts without altering granulocyte production or survival. Moreover, corticosteroid doses in the range used for other autoimmune disorders are sufficient to cause complications including increased susceptibility to infection. The poor response to corticosteroid therapy in many patients described in the early literature probably relates to poorly defined autoimmune neutropenia. In vitro evidence of antigranulocyte autoimmunity should be documented prior to the use of such therapeutic agents. Curiously, results from our laboratory suggest that, by virtue of their antilymphocyte activity, corticosteroids may be quite useful in treating the granulocytopenia associated with ADLG (30).

Other immunosuppressive agents, such as cyclophosphamide and azathioprine, may play a therapeutic role in selected patients (3, 11). By removing immune complexes and IgG, large volume plasmapheresis may also be beneficial in the treatment of autoimmune neutropenia.

SUMMARY

Because of the difficulties in working with granulocytes in vitro, elucidation of autoimmune neutropenic syndromes has been slower than has been the case with autoimmune hemolytic anemia and thrombocytopenia. With the development of sophisticated techniques for the study of granulocyte-associated antibodies, they have been found in disorders associated with neutropenia with increasing frequency; they appear to play a role in the pathogenesis of neutropenia in at least some of these disorders. Moreover, these assays have a role in planning treatment of the patient with suspected autoimmune neutropenia.

ACKNOWLEDGMENTS

This work was supported by the Research and Education branch of the Durham VA Hospital and by the Southeastern Cancer Therapy Group (Public Health service grant CA05634).

Literature Cited

1. Lalezari, P., Murphy, G. B., Allen, F. H. Jr. 1971. NB1, a new neutrophil-specific antigen involved in the pathogenesis of neonatal neutropenia. *J. Clin. Invest.* 50:1108
2. Weitzman, S. A., Stossel, T. P. 1978. Drug-induced immunological neutropenia. *Lancet* 1:1068
3. Starkebaum, G., Price, T. H., Lee, M. Y., Arend, W. P. 1978. Autoimmune neutropenia in systemic lupus erythematosus. *Arthritis Rheum.* 21:504
4. Boxer, L. A., Greenberg, M. S., Boxer, G. J., Stossel, T. P. 1975. Autoimmune neutropenia. *N. Engl. J. Med.* 293:748
5. Lalezari, P., Jiang, A., Yegen, L., Santorineou, M. 1975. Chronic autoimmune neutropenia due to anti-NA2 antibody. *N. Engl. J. Med.* 293:744
6. Craddock, P. R., Hammerschmidt, D., White, J. G., Dalmasso, A. P., Jacob, H. S. 1977. Complement (C5a)-induced granulocyte aggregation in vitro. *J. Clin. Invest.* 60:260
7. Logue, G. 1976. Felty's syndrome: Granulocyte-bound immunoglobulin G and splenectomy. *Ann. Intern. Med.* 85:437
8. Logue, G. L., Silberman, H. R. 1979. Felty's syndrome without splenomegaly. *Am. J. Med.* 66:703
9. Rosenthal, F. D., Beeley, J. M., Gelsthorpe, K., Doughty, R. W. 1974. White-cell antibodies and the aetiology of Felty's syndrome. *Q. J. Med.* 170:187
10. Calabresi, P., Edwards, E. A., Schilling, R. F. 1979. Fluorescent antiglobulin studies in leukopenic and related disorders. *J. Clin. Invest.* 38:2091
11. Verheugt, F. W. A., von dem Borne, A. E. G., van Noord-Bokhorst, J. C., Engelfriet, C. P. 1978. Autoimmune granulocytopenia: the detection of granulocyte autoantibodies with the immunofluorescence test. *Br. J. Haematol.* 39:339
12. Boxer, L. A., Stossel, T. P. 1974. Effects of anti-human neutrophil antibodies in vitro. *J. Clin. Invest.* 53:1534
13. McCallister, J. A., Boxer, L. A., Boehner, R. L. 1979. Detection of anti-polymorphonuclear leukocyte antibodies utilizing labeled staphylococcal Protein A. *Blood* In press
14. Williams, R. C. Jr., Bankhurst, A. D., Montavo, J. D. 1976. IgG antilymphocyte antibodies in SLE detected by ^{125}I Protein A. *Arthritis Rheum.* 19:1261
15. Drew, S. I., Carter, B. M., Guidera, D. P., Terasaki, P. I. 1977. Autoimmune cytotoxic granulocyte antibodies in health and disease. *Transplant. Proc.* IX:1887
16. Pruzanski, W., Armstrong, M., Urowitz, M. B. 1978. Heterogeneity of cold- and warm-reacting cytotoxins against lymphocytes, granulocytes, and monocytes in rheumatic diseases. *Clin. Immunol. Immunopathol.* 11:142
17. Logue, G., Kurlander, R., Pepe, R., Davis, W., Silberman, H. 1978. Antibody-dependent lymphocyte-mediated granulocyte cytotoxicity in man. *Blood* 51:97
18. Logue, G. L. 1978. Antibody-dependent cell-mediated granulocytotoxicity: a mechanism of failure of splenectomy therapy in Felty's syndrome. *Blood* 52(Suppl. 1):153
19. Felty, A. R. 1924. Chronic arthritis in the adult, associated with splenomegaly and leukopenia. *Bull. Johns Hopkins Hosp.* 35:16
20. Barnes, C. G., Turnbull, A. L., Vernon-Roberts, B. 1971. Felty's syndrome. A clinical and pathological survey of 21 patients and their response to treatment. *Ann. Rheum. Dis.* 30:359
21. Pengelly, D. D. R. 1966. Felty's syndrome. Good response to adrenocorticosteroids, possible mechanism of the anemia. *Br. Med. J.* 2:986
22. Laszlo, J., Jones, R., Silberman, H. R., Banks, P. M. 1978. Splenectomy for Felty's syndrome. Clinicopathological study of 27 patients. *Arch. Intern. Med.* 138:597
23. Bishop, C. R., Rothstein, G., Ashenbrucker, H. E., Athens, J. W. 1971. Leukokinetic studies. 14. Blood neutrophil kinetics in chronic steady state neutropenia. *J. Clin. Invest.* 50:1678
24. Vincent, P. C., Levi, J. A., MacQueen, A. 1974. The mechanism of neutropenia in Felty's syndrome. *Br. J. Haematol.* 27:463
25. Abdou, N. I., NaPombejara, C., Balentine, L., Abdou, N. L. 1978. Suppressor cell-mediated neutropenia in Felty's syndrome. *J. Clin. Invest.* 61:738
26. Dukham, D. J., Rhyne, R. L. Jr., Smith, F. E., Williams, R. C. Jr. 1975. Retardation of colony growth of *in vitro* bone marrow culture using sera from patients with Felty's syndrome, SLE, rheumatoid arthritis and other disease states. *Arthritis Rheum.* 18:323
27. Budman, D. R., Steinberg, A. B. 1977. Hematologic aspects of systemic lupus erythematosus. *Ann. Intern. Med.* 86: 220

28. Dausset, J., Colombani, J., Colombani, M. 1961. Study of leukopenias and thrombocytopenias by the direct antiglobulin consumption test on leukocytes and/or platelets. *Blood* 18:672
29. Winfield, J. B., Winchester, R. J., Wernet, P., Fu, S. M., Kunkel, H. G. 1975. Nature of cold-reactive antibodies to lymphocyte determinants in systemic lupus erythematosus. *Arthritis Rheum.* 18:1
30. Logue, G. L. Unpublished observations
31. Evans, R. S., Duane, R. T. 1949. Acquired hemolytic anemia. I. the relation of erythrocyte antibody production to activity of the disease. II. the significance of thrombocytopenia and leukopenia. *Blood* 4:1196
32. Crosby, W. H. 1973. What to treat in Felty's syndrome. *J. Am. Med. Assoc.* 225:1114
33. Sienknecht, C. W., Urowitz, M. B., Pruzanski, W., Stein, H. B. 1977. Felty's syndrome. Clinical and serological analysis of 34 cases. *Ann. Rheum. Dis.* 36:500
34. Rivero, S., Alger, M., Alarcin-Seguria, D. 1979. Splenectomy for hematocytopenia in systemic lupus erythematosus. *Arch. Intern. Med.* 139:773

Ann. Rev. Med. 1980. 31:201–18
Copyright © 1980 by Annual Reviews Inc. All rights reserved

NONINVASIVE EVALUATION OF THE CAROTID BRUIT

❖7354

Robert W. Barnes, M.D.[1]

Department of Surgery, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia 23298; and McGuire Veterans Administration Medical Center, Richmond, Virginia 23249

INTRODUCTION

The most common clinical sign of extracranial carotid occlusive disease is the presence of a carotid bruit audible with a stethoscope. The bruit is the result of arterial wall vibrations induced by disturbed or turbulent blood flow immediately distal to a stenosis (narrowing), which reduces the vascular lumen by 30–85%. Occasionally a bruit will result from the increased blood flow through a normal external carotid artery in the presence of internal carotid artery occlusion. Rarely bruits may occur in high flow states in the absence of carotid artery disease, as in anemia or hyperthyroidism.

Two errors are often made in evaluating patients for carotid bruits. One is to mistake a bruit arising from the subclavian, innominate, or common carotid artery, or the aortic valve, for a carotid bifurcation bruit. Such radiating thoracic vascular bruits account for approximately 50% of all cervical bruits. Their proper identification is important inasmuch as they are not as frequently associated with cerebral ischemic episodes as are true carotid bruits. Subclavian artery bruits are loudest in the supraclavicular areas and common carotid bruits are loudest low in the anterior neck. A bruit over the right subclavian and common carotid arteries suggests innominate (brachiocephalic) artery stenosis. A carotid bifurcation bruit is heard maximally over the midportion or the upper area of the neck, between the superior border of the thyroid cartilage and the angle of the mandible.

---

[1] Supported in part by NIH Grant 1 RO1 HL 23568–01 and research funds from the Veterans Administration.

0066-4219/80/0401-0201$01.00

The second mistake is to equate a carotid bruit with hemodynamically significant stenosis of the internal carotid artery. Judged by noninvasive diagnostic screening techniques that are sensitive to flow-reducing lesions, only about 20% of carotid bruits are of hemodynamic significance, i.e. associated with ≥ 50% diameter reduction of the internal carotid artery relative to the distal uninvolved vessel (70–80% area reduction). Furthermore a carotid bruit is both insensitive and nonspecific with respect to internal carotid artery disease. Stenosis or ulcerated plaques with less than 30% lumen compromise do not result in a bruit, and stenosis of greater than 85% may reduce flow to the point that a bruit becomes inaudible. Internal carotid occlusion often is not associated with a bruit. Finally a carotid bifurcation bruit may reflect isolated stenosis of the external carotid artery. The only characteristic quality of a carotid bruit that is pathognomonic of severe internal carotid artery stenosis is a bruit that extends throughout systole and into diastole, as a result of the high diastolic flow associated with low cerebrovascular resistance.

A carotid bruit in a patient without cerebrovascular symptoms, termed "asymptomatic," is sometimes discovered incidently. The evaluation of such a patient is important because about 50% of patients with cerebral transient ischemic attack (TIA) or stroke have extracranial cerebrovascular disease which is surgically accessible. Inasmuch as carotid disease may occur without a bruit, it may be more appropriate to identify the stroke-prone patient as an individual with "asymptomatic carotid disease" rather than an asymptomatic carotid bruit.

Contrast arteriography remains the "gold standard" for diagnosis of cerebrovascular disease. Although this invasive procedure is appropriate for patients with symptoms of TIA or stroke, the risk and expense of arteriography preclude its routine use as a screening and follow-up procedure. The assessment of asymptomatic patients or those with atypical and nonlocalizing symptoms has been greatly enhanced by the development of several noninvasive cerebrovascular diagnostic techniques during the past decade. This chapter reviews the principles, clinical applications, attributes, and limitations of the more common noninvasive screening methods, and suggests some future developments in this field.

NONINVASIVE TECHNIQUES

Noninvasive cerebrovascular screening techniques may be either indirect or direct. Indirect techniques assess alterations of pressure, flow, or pulse arrival in the ophthalmic artery (a branch of the internal carotid artery) or its branches. Such techniques may detect hemodynamically significant stenosis (≥ 50% diameter reduction) or occlusion of the internal carotid artery. Indirect methods include periorbital Doppler ultrasound, supraorbital or ocular plethysmography, ophthalmodynamometry, tonography, and facial thermography. Direct techniques assess bruits, flow, or morphology of the carotid artery and its branches in the neck. Such techniques include carotid phonoangiography, carotid flow velocity analysis by Doppler ultrasound, carotid imaging by Doppler and/or B-mode ultrasound, and dynamic radionuclide scanning.

Indirect Methods

PERIORBITAL DOPPLER ULTRASOUND The Doppler ultrasonic velocity detector contains a piezoelectric crystal in a hand-held probe that emits ultrasound at a frequency of 5–10 MHz. The ultrasound is transmitted into the tissues via acoustic gel coupling with the skin. Sound reflected from moving blood cells is shifted in frequency by an amount proportional to their velocity. The backscattered ultrasound is received by a second crystal and the frequency shift is detected and amplified by the instrument as an audible signal or a recordable analogue waveform. Direction-sensing Doppler units are capable of determining whether blood is flowing toward or away from the Doppler probe.

The periorbital Doppler examination involves the determination of blood flow direction in the ophthalmic artery branches (frontal and supraorbital arteries) and the effect of transient compression of external carotid branches (superficial temporal, infraorbital, and facial arteries) and each common carotid artery (1–3). Ophthalmic artery flow is normally antegrade out of the orbit; the flow is either unaffected or augmented by external carotid branch compression and diminished only by ipsilateral common carotid compression. With severe stenosis ($\geqslant 50\%$) or occlusion of the extracranial internal carotid artery, ophthalmic artery flow may be retrograde and diminished by compression of one of the external carotid branches. Alternately, collateral flow from the opposite internal carotid artery via the circle of Willis may be identified by normal directional ophthalmic artery flow, which is attenuated only by contralateral common carotid compression. If compression of each carotid artery fails to attenuate ophthalmic artery flow, severe internal carotid obstruction with collateral via the vertebrobasilar system is inferred.

With experience, periorbital Doppler examinations may detect significant stenosis ($\geqslant 50\%$) or occlusion of the internal carotid artery with a sensitivity and specificity exceeding 95% (4) (Table 1). Doppler ultrasound is the least expensive of the noninvasive techniques and has the added attributes of simplicity, rapidity, small size, and portability. This instrument also permits direct carotid interrogation of flow velocity (see below) and

Table 1 Indirect cerebrovascular screening techniques[a]

Technique	Simplicity	Objectivity	Inexpense	Portability	Sensitivity	Specificity
Periorbital Doppler	++	+	++	++	++	++
Supraorbital photoplethysmography	++	++	+	+	++	+
Ocular plethysmography	+	++	0	+	++	++
Ophthalmodynamometry	+	++	++	++	+	+
Carotid compression tonography	+	++	+	+	++	+
Thermography	0	++	0	0	+	+

[a] ++ = good; + = fair; 0 = poor.

forms the backbone of many noninvasive vascular laboratories because of its added versatility in peripheral arterial and venous disease. However, periorbital Doppler examination requires considerable experience to achieve maximal accuracy, and like the indirect techniques, it is not sensitive to nonobstructing (< 50% stenosis) or ulcerated carotid plaques, which may be an important cause of cerebral emboli (5).

SUPRAORBITAL PHOTOPLETHYSMOGRAPHY The photoelectric plethysmograph (PPG), while available for over 40 years, has recently been improved to include an infrared light-emitting diode and phototransistor in the tranducer. The device permits sensitive detection of pulsatile changes in blood content of the skin (ac mode) and some types also determine relative skin blood content (dc mode); the latter are useful in assessment of chronic venous disease. To screen for cerebrovascular disease (6), a PPG transducer is affixed to the skin above the medial end of each eyebrow. The transducers are connected via an amplifier to a two-channel strip chart recorder to obtain pulsatile tracings of the supraorbital microcirculation. This area of the forehead normally receives its circulation from the supraorbital and frontal artery branches of the ophthalmic artery. Normally the supraorbital pulse amplitude is not significantly attenuated by compression of external carotid branches but decreases only with ipsilateral common carotid compression. If attenuation of supraorbital PPG pulse amplitude occurs with the other compression maneuvers, significant internal carotid obstruction is inferred.

Supraorbital PPG has a sensitivity of over 95% in detection of hemodynamically significant internal carotid obstruction, but there are frequent false positive results so that the specificity is only 80–90%. The method is simple, rapid, portable, and relatively inexpensive. It provides a hard-copy record of the simultaneous bilateral examination. Unlike ocular plethysmography (OPG), the method does not require anesthesia of the eyes, and the accuracy of the test is not influenced by the presence of bilateral carotid disease of equal hemodynamic severity. Supraorbital PPG's greatest limitation is its high incidence of false positive results. An abnormal study should prompt examination by another technique, such as periorbital Doppler examination. When the two techniques agree, the accuracy of detection or exclusion of hemodynamically significant carotid occlusive disease exceeds 95% (7). Obviously, the method does not detect nonobstructive or ulcerated carotid plaques.

OCULAR PLETHYSMOGRAPHY Pressure transducers held by suction on the anesthetized eye can detect ocular pulse delays or reduced retinal artery pressure in the presence of severe carotid occlusive disease. Pulse delay methods include the water-filled ocular plethysmograph (OPG) of Kartchner & McRae (8) and the air-filled digitized system of Zira (9). These techniques detect both amplitude reduction and ocular filling-pulse delays relative to the opposite eye in unilateral carotid disease. In bilateral carotid disease both ocular pulse delays may be detected relative to the ear pulse, the latter being determined by photoplethysmography. Retinal artery pressure is determined indirectly by the oculopneumoplethysmograph of Gee et al (10) (OPG-Gee), which records the return of ocular pulsation relative to the recorded suction negative pressure on the transducer as it is slowly reduced from –300 mm Hg (–500 mm Hg in hypertensive patients). The technique also permits estimation of hemispheric collateral perfusion pressure from the retinal artery pressure determined during common carotid compression (11). Ocular pulse delays relative to the electrocardiogram may be measured with the OPG-Gee.

The sensitivity and specificity of OPG are generally reported to be from 85 to 95%, although some investigators have noted much lower accuracy (12). The techniques are often used in conjunction with carotid phonoangiography (CPA), which provides complementary information especially in the presence of bilateral carotid occlusive disease (8). The advantages of the techniques include their objectivity, rapidity, and the provision of a hard-copy record. An added advantage of the OPG-Gee is its capability of estimating hemispheric collateral perfusion pressure, which may help predict the need for selective shunting during carotid endarterectomy. Such information may also indicate the potential risk of stroke if the carotid artery becomes occluded or needs to be sacrificed during neck dissection for tumor. Disadvantages of OPG include its relatively higher cost and its occasional insensitivity to bilateral carotid occlusive disease of balanced hemodynamic severity. The techniques require ocular anesthesia and some degree of patient cooperation. The techniques cannot be performed in a patient with significant ocular disease or a missing eye. Ocular injuries as a complication of the techniques (corneal abrasion, subconjunctival hemorrhage) are rare. The OPG methods are not sensitive to nonobstructive or ulcerated carotid plaques.

OPHTHALMODYNAMOMETRY Retinal artery pressure may be indirectly assessed by measuring the force applied on the anesthetized sclera by the footplate of the ophthalmodynamometer (ODM) (13). Using an ophthalmoscope, the systolic value is noted at that point when the retinal arteries cease to pulsate, and the diastolic pressure is determined at that point when the arteries develop maximal pulsation. The technique, while inexpensive, simple, and portable, is less accurate (75–85%) than other methods and is insensitive to bilateral severe carotid disease as well as nonobstructive or ulcerated plaques.

CAROTID COMPRESSION TONOGRAPHY Carotid compression tonography uses a recording tonometer to measure ocular pressure during transient common carotid compression (14). Normally the ocular pressure falls rapidly during ipsilateral carotid compression and recovers promptly after release of compression. In the presence of significant carotid obstruction, the pressure fall may be attenuated or absent during ipsilateral carotid compression and may drop with contralateral carotid compression. If neither carotid compression alters the ocular pressure recording, collateral circulation from the vertebrobasilar system is inferred. Carotid compression tonography is fairly sensitive (92%) but has many false positives (75% specificity). The method also is insensitive to nonobstructive or ulcerated plaques.

THERMOGRAPHY Electronic monitoring of infrared emission from the body permits sensitive detection of relative differences in temperature of the surface of the body. Such temperature mapping of the face locates the normally warmer regions of each supraorbital area, which is supplied by the ophthalmic artery branches carrying blood at core temperature (15). In the presence of significant carotid occlusive disease and reversed ophthalmic artery flow via external carotid artery collaterals, the supraorbital area appears as a relatively cooler area. This technique will not detect disease associated with intracranial collateral circulation via the circle of Willis, and it is insensitive to nonobstructive carotid disease. Its primary drawback is the expense and complexity of the equipment.

Direct Methods

CAROTID PHONOANGIOGRAPHY The amplitude and duration of a carotid bruit may be recorded with a carotid phonoangiograph (CPA), which consists of a sensitive electronic microphone and an oscilloscope that permits a Polaroid® photograph of three separate areas of carotid interrogation (8). A short systolic bruit may be recorded from a stenosis of 30–40% diameter reduction. More severe stenosis may result in a pansystolic bruit of greater amplitude. Such bruits often arise from an internal carotid stenosis but may also be produced by a lesion in the common or external carotid artery. The only pathognomonic sign of severe (70–85%) stenosis of the internal carotid artery is a bruit that extends throughout systole and into diastole. In the presence of >85% stenosis or occlusion of the internal carotid artery, a bruit may not be detectable. The CPA is thus relatively insensitive although specific (usually normal in normal carotids) (Table 2). It is usually used only in conjunction with other indirect techniques, especially the OPG (8).

Recently the frequency content of the sound spectrum of a carotid bruit was analyzed by fast Fourier (16) and autocorrelation (17) techniques to more accurately estimate the severity of carotid stenosis. While these techniques may prove to be a more sensitive guide to the presence and degree of internal carotid stenosis, they are more expensive and remain insensitive to carotid disease that does not produce a bruit (<30% stenosis, or occlusion).

CAROTID FLOW ANALYSIS (DOPPLER) Blood flow velocity in the carotid artery and its branches may be directly measured by Doppler ultrasound. The flow signals may be assessed by audible interpretation (18), analogue recordings (19), or sound spectral analysis of the frequency and

Table 2 Direct cerebrovascular screening techniques[a]

Technique	Simplicity	Objectivity	Inexpense	Portability	Sensitivity	Specificity
Phonoangiography	++	++	+	++	0	++
Carotid flow analysis	++	+	++	++	++	++
Doppler imaging	+	+	0	+	++	++
B-scanning	+	++	0	0	++	++
Duplex scanning	+	++	0	0	++	++
Radionuclide anteriography	+	+	0	0	0	++

[a] ++ = good; + = fair; 0 = poor.

amplitude content of the Doppler signal (20). The velocity sampling may be from across the entire vascular lumen, using conventional continuous-wave devices, or from discrete sample volumes within the vessel, obtained with a pulsed-Doppler unit. Using discriminant analysis of multiple dimensions from analogue recordings of common carotid artery velocity, Rutherford et al reported 95% accuracy compared to arteriography distinguishing young and old "normal" carotids, <50% stenosis, >50% stenosis, and occlusion (19). All normal, stenotic, and occluded arteries were correctly distinguished, which was not possible by periorbital Doppler examination.

With experience, stenotic or occlusive disease in carotid artery segments can be detected and localized by using simple audible assessment of continuous-wave Doppler signals; the accuracy exceeds 90% (18). More sensitive and objective assessment of carotid artery disease, including detection of stenosis as insignificant as 10–25%, has been achieved using sound spectral analysis of the frequency signal (20). Recent development of real-time spectral analysis may greatly increase the accuracy of identifying carotid artery disease. Such techniques are still investigational and relatively expensive. Spectral analysis will probably become an integral part of Doppler ultrasonic imaging systems.

DOPPLER ULTRASONIC ARTERIOGRAPHY Noninvasive imaging of the carotid artery flow lumen is feasible using either pulsed (21) or continuous-wave (22) Doppler instruments. These devices consist of the Doppler probe, which is constrained to move in a position-sensing arm so that the location of detected flow may be appropriately positioned on a storage oscilloscope. The carotid artery lumen is progressively imaged as the Doppler probe is moved over the course of the vessel. The resultant image may be photographed. Views are possible not only in the projection mode, similar to conventional contrast arteriograms, but also in unique transverse or longitudinal cross-sectional images. The instruments also permit flow velocity sampling for audible, analogue, or sound spectral analysis, which enhances the diagnostic interpretation (23). Such flow analysis is particularly important to detect stenosis in arterial segments that cannot be imaged satisfactorily because of vascular wall calcification.

The sensitivity and specificity of Doppler ultrasonic arteriography range from 80 to 90%; it detects stenoses of <50% with much greater frequency than do indirect periorbital techniques. The technique also permits discrimination of stenosis from occlusion of the internal carotid artery. Although vascular wall calcification may impair carotid imaging in as many as 50% of patients, concomitant audio or spectral analysis of carotid flow velocity signals may improve accuracy to above 90%. The drawbacks to ultrasonic arteriography include the expense and the technical expertise required for maximal accuracy in use of the equipment.

B-MODE ULTRASONIC SCANNING Real-time imaging of the carotid artery may be carried out by B-mode ultrasound (24). By using 5-MHz pulsed ultrasound, echos from the carotid vascular wall and associated plaques are stored in real time on a grey-scale oscilloscope or on a video screen. The instruments have been progressively refined to provide resolution of 1.0 mm or less and to discriminate such vascular wall constituents as fatty plaque, fibrous tissue, calcium, or blood. The merits of B-scanning include rapid visualization, increased sensitivity to nonobstructive or ulcerated plaques, and capability of measuring and following the size and evolution of atherosclerotic lesions. Disadvantages include the high cost, limited portability, and limitations in discriminating advanced stenosis from occlusion. Calcification of the vascular wall may also cause acoustic shadowing that precludes definition of the subjacent vascular wall or lumen. Finally, early fatty plaques have the acoustic impedance of blood and may not be detected by B-mode scanning. Validity of B-scan equipment is currently under investigation.

DUPLEX ULTRASONIC IMAGING Real-time B-scanning and Doppler ultrasonic assessment of carotid flow velocity may be combined in the duplex ultrasonic imager (25). The lumen of the B-mode imaged carotid segments may be interrogated by moving the pulsed-Doppler sample volume across the flow stream. The resultant flow signal may be assessed audibly or by sound spectral analysis of the frequency. Preliminary studies suggest that this technique is sensitive to carotid stenosis of 10–20%, and the method is capable of distinguishing advanced stenosis from occlusion, which may not be possible with B-scanning alone. Validation of the technique is being initiated. The attributes of this technique are the combined advantages of B-mode scanning and pulsed-Doppler ultrasound. The disadvantages are its expense, relative lack of portability, and technical complexity.

RADIONUCLIDE ARTERIOGRAPHY Dynamic scanning of the carotid artery and the cerebral flow phase of the radionuclide brain scan permits assessment of severe carotid occlusive disease (26). Intravenous injection of technetium allows imaging of the cerebrovascular flow prior to the static brain scan. Although the technique improves the diagnostic information of brain scanning, the sensitivity and specificity seldom exceed 80%, and the technique is not usually as reliable as other noninvasive techniques for screening of carotid occlusive disease.

CLINICAL APPLICATIONS

Patients with suspected cerebrovascular disease may be classified in one of these groups: (*a*) those with localizing symptoms such as unilateral motor or sensory disturbance, monocular blindness (amaurosis fugax), or dysphasia; (*b*) patients with nonspecific (nonlateralizing) symptoms such as dizziness, syncope, headache, bilateral visual disturbance, confusion, or memory disturbance; and (*c*) asymptomatic patients.

Patients with clinical hemispheric (carotid territory) or posterior (vertebrobasilar) symptoms, whether the result of TIA's or stroke, should probably have contrast arteriography to determine whether a surgically correctable lesion is present. The role of noninvasive cerebrovascular screening in such patients is limited, since over 50% of patients who are surgical candidates have a normal noninvasive examination, and presumably their symptoms arise from emboli from nonobstructive (<50% stenosis) or ulcerated carotid plaques (27). Noninvasive techniques may be used for perioperative monitoring of these patients during and after carotid endarterectomy (28). These techniques also permit longitudinal follow-up of the natural history and efficacy of medical or surgical therapy.

Patients with nonspecific or nonlocalizing symptoms are often classified with or managed as asymptomatic patients. However, patients with nonhemispheric symptoms may benefit from carotid endarterectomy, with relief of symptoms reported in 67% and improvement in 28% (29). Such improvement may not correlate with the results of noninvasive periorbital Doppler examination (30). The primary clinical application of noninvasive cerebrovascular screening techniques has been the assessment of patients with an asymptomatic carotid bruit or the risk of harboring asymptomatic carotid disease, regardless of the presence or absence of a bruit. The remainder of this review addresses the problem of asymptomatic carotid disease that may present in one of three clinical circumstances: (*a*) as isolated carotid disease; (*b*) with concomitant major operation; and (*c*) with contralateral symptomatic carotid disease.

Isolated Asymptomatic Carotid Bruit

Four studies of patients with asymptomatic carotid bruit, with or without prophylactic carotid endarterectomy, are summarized in Table 3. No randomized studies comparing nonoperative therapy with prophylactic carotid endarterectomy have been reported.

Javid et al (31) reported 56 patients with asymptomatic carotid bruit and associated hypertension or anticipated major general surgical procedures. All patients underwent prophylactic carotid endarterectomy with one operative death and two postoperative strokes. There were two late strokes

Table 3 Isolated asymptomatic carotid bruit

				Nonfatal		Deaths	
Author (year)	Follow-up period (yr)	No. of patients	Treatment	TIA	Stroke	Stroke-related	Other
Javid et al (1971)	1–6	56	Endarterectomy	—	2	0	15
Thompson et al (1978)	1–20	132	Endarterectomy	6	3	3	40
		138	No operation	37	24	3	47
Cooperman et al (1978)	2–7	60	Bruit—postop. vascular	12	5	4	21
		196	No bruit—reconstruction	23	6	1	31
Moore et al (1978)	1–8	38	Nonoperated shallow ulcers	—	0.4%/yr	—	53%/8 yr
		29	Nonoperated deep ulcers	—	12.5%/yr	—	79%/8 yr
		68	Operated symptomatic ulcer	—	1.5%/yr	—	55%/8 yr

during the follow-up period of up to six years. Of 15 late deaths, 12 resulted from myocardial infarction, and none were caused by stroke.

Thompson et al (32) reported the largest series of patients with asymptomatic carotid bruit managed either nonoperatively or by prophylactic carotid endarterectomy. The incidence of late (up to 20 years) TIA or stroke in 132 patients undergoing operation, 4.5% and 4.6% respectively, was significantly lower than the incidence in 138 nonoperated patients, 26.8% and 17.4%, respectively. However the incidence of stroke-related deaths and the total number of late deaths was similar in the two groups.

Cooperman et al (33) noted a significantly greater incidence of TIA (20% vs 12%), stroke (15% vs 4%), stroke-related death (7% vs 0.6%), and total late deaths (42% vs 16%) in patients with bruits compared with those without bruits up to seven years following major peripheral arterial reconstruction.

Moore et al (34) recently reported on the increased risk of late stroke (12.5% per year) in asymptomatic patients with significant carotid ulceration as compared to those with insignificant ulceration (0.4% per year) or patients undergoing endarterectomy for symptomatic carotid ulceration (1.5% per year).

Although noninvasive cerebrovascular screening has been recommended for patients with asymptomatic carotid bruits, the only large study to date with reported follow-up of the incidence of stroke and/or endarterectomy in such patients is that of Kartchner & McRae (35; Table 4). Of more than 4000 patients evaluated with ocular plethysmography (OPG) and carotid phonoangiography (CPA), 1287 were studied for "asymptomatic" carotid

Table 4 Noninvasive screening of asymptomatic bruit (35)

CPA[a]	OPG[b]	No. of patients[c]	TIA	Stroke	Endarterectomy	Deaths
Negative	Negative	755	16	14	15	63
Negative	Positive	130	5	4	29	10
Positive	Negative	102	3	3	13	12
Positive	Positive	143	7	17	78	17

[a] CPA = carotid phonoangiography.
[b] OPG = ocular plethysmography.
[c] Follow-up period 6–70 months, average 24 months.

bruits. Of these, 697 had nonhemispheric neurological symptoms and 590 were clinically asymptomatic. Radiating subbifurcation bruits were diagnosed by CPA in 619, "insignificant" bifurcation bruits in 426, and significant carotid bruits in 242 patients. By OPG, 273 patients (21%) had hemodynamically significant obstruction of the internal carotid artery and 5% had external carotid stenosis.

Of the 1287 patients, 1130 (88%) were followed for periods ranging from 6 to 70 months with an average of 24 months. In the absence of a significant carotid bruit by CPA and a negative OPG examination, only 2% of 755 patients came to carotid endarterectomy and the incidence of TIA and stroke in the follow-up period was 2.1% and 1.9%, respectively. Of 143 patients with a significant carotid bruit and an abnormal OPG examination, 54.6% came to carotid endarterectomy and the incidence of TIA and stroke were 4.9% and 11.9%, respectively, with only one stroke occurring in relation to a carotid endarterectomy. Although the study does not deal with patients randomized by type of therapy, the data tend to support a correlation between the hemodynamic severity of carotid stenosis, determined noninvasively, and the subsequent development of symptoms of cerebral ischemia.

Concomitant Major Operation

Although the role of prophylactic carotid endarterectomy in a patient with isolated asymptomatic carotid bruit is unknown, this procedure is frequently recommended for the patient who is to undergo another major operation. However, the stroke risk of such patients is not clear and several reports (36–42) suggest that an asymptomatic bruit does not correlate with subsequent stroke after major operation (Table 5). The presumed cause of stroke in such patients is thrombosis of a severely stenotic carotid artery during a period of hypotension or low flow in the perioperative period. The two major operative procedures with the greatest prevalence of associated carotid artery disease are coronary artery bypass and major peripheral

Table 5 Asymptomatic bruit and concomitant major operation

Author (year)	No. of patients	Study material	Post endarterectomy Stroke	Post endarterectomy Death	Post major operation Stroke	Post major operation Death
Treiman	40	With bruit—elective	—		0	
et al (1973)	157	Without bruit—aortic operation	—	8	2	8
Lefrak & Guinn (1974)	34	Preop. major surgery (esp. vascular)	1	1	0	5
Carney et al	35	Bruit—preop.	—	—	0	
(1977)	139	No bruit—aortoiliac	—	—	2	2
	74	?No bruit—reconstruction	—	—	2	
Evans &	92	Bruit—preop. vascular	—	—	0	—
Cooperman (1978)	496	No bruit—reconstruction	—	—	4	—
Bernhard	15	Staged carotid-coronary	0	3[a]	0	2
et al (1972)	15	Simultaneous carotid-coronary	1	0	—	—
Hertzer	59	Staged	2	0	0	1
et al (1978)	115	Simultaneous	10	5	—	—
Ennix et al	77	Staged	4	14	—	—
(1979)	135	Antecedent (84) or simultaneous (51) coronary bypass	2	4	—	—

[a] Myocardial infarction.

arterial reconstruction. The fact that cardiac or vascular surgeons also perform carotid endarterectomy may explain the frequent tendency for such surgeons to correct significant carotid artery disease prior to the intended operations. However, several of the reported studies have documented a low incidence of perioperative stroke in patients who did not undergo prophylactic carotid endarterectomy. Most of the studies suggesting a nonoperative approach to asymptomatic carotid bruit prior to major operation involve peripheral arterial reconstruction. The reports involving coronary artery bypass usually recommend prophylactic carotid artery endarterectomy either prior to the coronary bypass or as a concomitant procedure.

All three series reporting perioperative stroke in relation to peripheral artery reconstruction noted the postoperative neurologic deficit in patients who specifically did not have a carotid bruit (36, 38, 39). One explanation for this dissociation of asymptomatic bruit and perioperative stroke is the fact that only a minority of so-called carotid bruits are associated with hemodynamically significant internal carotid stenosis. As noted in Kartchner & McRae's study (35), nearly 50% of patients with "carotid" bruit have subbifurcation bruits radiating from the great vessels of the thorax. Those authors noted only a 21% incidence of significant carotid obstruction by OPG in patients screened for asymptomatic carotid bruits. In our laboratory the incidence of hemodynamically significant carotid occlusive disease in patients with asymptomatic bruit has been 16%. Another factor explaining the lack of correlation of asymptomatic bruit and perioperative stroke may be the fact that preocclusive carotid stenosis of >85% often is not associated with a bruit. Thus the patient who may be most susceptible to perioperative carotid thrombosis may have a stenosis which is not evident by audible bruit. Such patients can only be detected by routine screening with noninvasive techniques.

Patients with concomitant asymptomatic carotid and symptomatic coronary artery disease have usually had prophylactic carotid endarterectomy prior to coronary artery bypass, either as a staged or a concomitant procedure (40–42). Bernhard et al (40) first suggested the merits of concomitant carotid and coronary reconstruction, because of a 20% incidence of fatal myocardial infarction prior to coronary bypass if the procedures were staged. Although many cardiovascular surgeons prefer simultaneous carotid and coronary procedures, Hertzler et al (41) found a higher incidence of perioperative stroke (4.3%) and death (4.3%) with concomitant operations compared to staged procedures (1.7% and 1.7% respectively). Most surgeons recommend carotid endarterectomy prior to coronary bypass, reserving simultaneous procedures for patients with severe coronary artery disease (unstable angina, left main coronary disease or its equivalent). However, the recent report by Ennix et al (42) documents the safety of combined carotid and coronary reconstruction in experienced hands.

Few studies have reported the role of noninvasive diagnostic techniques in selecting patients with asymptomatic carotid disease for angiography and prophylactic endarterectomy prior to major operation. Some of the patients in the study by Javid et al (31) were preoperative patients. Brewster et al (43) reported 57 patients undergoing several noninvasive screening techniques (periorbital Doppler, OPG, CPA) prior to major operation. All 45 patients with negative screening studies were operated without postoperative neurologic deficit. Eight of 12 patients with abnormal carotid screening studies had angiography and prophylactic carotid endarterectomy with no subsequent neurologic deficit after the intended operation. Four patients did not undergo arteriography and one suffered a stroke following the major operation. In a pilot study of 50 general surgical patients screened preoperatively by periorbital Doppler ultrasound, we noted one stroke following a colectomy in one of three patients with reversed opthalmic artery flow. This study led us to embark on a three-year prospective study to screen noninvasively for carotid disease in all patients undergoing coronary or peripheral artery reconstruction and all general surgical patients with an asymptomatic bruit. Patients with abnormal carotid studies will undergo arteriography and those with operable severe carotid stenosis will be randomly allocated to a prophylactic endarterectomy or a control group. This will be the first major randomized study to assess the efficacy of prophylactic carotid endarterectomy in patients undergoing major operation.

Contralateral Symptomatic Carotid Disease

The identification of an asymptomatic carotid lesion opposite symptomatic carotid disease on contrast arteriography often leads to a therapeutic dilemma. Many vascular surgeons perform staged carotid endarterectomies, assuming that the risk of prophylactic repair of the asymptomatic lesion is less than the future risk of stroke, especially since the risk of arteriography has already passed. However, several studies (44–46) suggest a low incidence of future stroke in asymptomatic carotid stenosis after endarterectomy of a contralateral symptomatic lesion (Table 6). That the stroke rate may be lower than that of patients with isolated asymptomatic carotid disease may relate to the improved cerebral hemodynamics associated with the endarterectomy of the symptomatic stenosis.

Humphries et al (45) emphasized that only one of 168 patients who were followed with asymptomatic carotid disease after contralateral endarterectomy developed a stroke without antecendent TIA. Three patients who refused operation after developing TIAs developed a subsequent stroke. Another 15.5% of patients developed TIA and underwent successful endarterectomy without developing stroke.

CONCLUSIONS

During the past decade a number of noninvasive cerebrovascular diagnostic techniques have been developed to screen patients for extracranial carotid occlusive disease. Because about 50% of patients with stroke develop cere-

Table 6 Asymptomatic bruit with contralateral symptomatic disease

Author (year)	Follow-up period (yr)	No. of patients	TIA	Stroke	Endarterectomy	Death
Levine & Sondheimer (1973)	2–12	60	2	0	0	2
Humphries et al (1976)	up to 12	168	26	4^a	26	—
Johnson et al (1978)	up to 4	103	3	0	3	7

[a] Antecedent TIAs preceded stroke in three patients who refused endarterectomy.

bral ischemia on the basis of extracranial carotid disease, the early detection of such lesions may lead to improved therapy and possible stroke prevention. However, further epidemiologic study is necessary to identify the natural history of asymptomatic carotid occlusive disease. Noninvasive screening techniques may permit more objective assessment of the prevalence and course of asymptomatic carotid disease, with or without a bruit.

Noninvasive screening techniques include indirect, or periorbital, and direct carotid surveillance methods. The former are less expensive, more rapid and portable, but they are insensitive to nonobstructive or ulcerated carotid plaques. Furthermore they cannot distinguish severe carotid stenosis, which is operable, from carotid occlusion, which is generally inoperable by extracranial techniques. Direct carotid screening techniques are more expensive and less portable but they are sensitive to a wider range of carotid disease, including nonobstructive stenosis, and they can differentiate carotid stenosis from occlusion.

Although a carotid bruit is the traditional sign of carotid stenosis, there are several limitations of this sign of cerebrovascular disease. Because bruits from the common carotid or subclavian artery or the aortic valve occur frequently, only about 50% of patients with cervical bruits have true carotid bifurcation bruits. Secondly, only about 20% of patients with "carotid bruits" have hemodynamically significant obstruction of the internal carotid artery. Finally, many patients with severe carotid occlusive disease do not have a bruit, including those with >85% stenosis and patients with occlusion of the internal carotid artery. Thus, to identify the "stroke-prone" patient, a clinician must be able to detect significant internal carotid occlusive disease regardless of the presence or absence of a carotid bruit. To meet this objective, the intelligent use of noninvasive cerebrovascular screening techniques has the greatest merit for patients with asymptomatic carotid artery disease.

The future of noninvasive cerebrovascular screening will probably encompass three broad areas: (*a*) improved instrumentation; (*b*) validation studies, both experimental and clinical; and (*c*) clinical application and outreach or transfer of the new technology to practicing physicians. Greater emphasis will be placed upon direct carotid screening techniques, particularly real-time B-mode ultrasonic scanning and Doppler flow velocity analysis to improve diagnostic sensitivity and discrimination of carotid stenosis from occlusions. Once validated, such duplex systems may not only improve screening of asymptomatic patients but also may permit more intelligent selection of angiography for patients who are candidates for endarterectomy of symptomatic stenoses, while excluding those patients who are not operative candidates (normal or occluded arteries), thus sparing them the risk and expense of contrast arteriography.

Literature Cited

1. Brockenbrough, D. C. 1969. *Screening for the Prevention of Stroke: Use of Doppler Flowmeter.* Brochure issued by Alaska/Washington Regional Medical Program
2. Müller, H. R. 1972. The diagnosis of internal carotid artery occlusion by directional Doppler sonography of the ophthalmic artery. *Neurology* 22: 816–23
3. Barnes, R. W., Wilson, M. R. 1975. *Doppler Ultrasonic Evaluation of Cerebrovascular Disease: A Programmed Audiovisual Instruction.* Iowa City: Univ. Iowa, 196 pp.
4. Barnes, R. W., Russell, H. E., Bone, G. E., Slaymaker, E. E. 1977. Doppler cerebrovascular examination: Improved results with refinements in technique. *Stroke* 8:468–71
5. Bone, G. E., Barnes, R. W. 1976. Limitations of the Doppler cerebrovascular examination in hemispheric cerebral ischemia. *Surgery* 79:577–80
6. Barnes, R. W., Clayton, J. M., Bone, G. E., Slaymaker, E. E., Reinertson, J. 1977. Supraorbital photoplethysmography. *J. Surg. Res.* 22:319–27
7. Barnes, R. W., Garrett, W. V., Slaymaker, E. E., Reinertson, J. E. 1977. Doppler ultrasound and supraorbital photoplethysmography for noninvasive screening of carotid occlusive disease. *Am. J. Surg.* 134:183–86
8. Kartchner, M. M., McRae, L. P., Morrison, F. D. 1973. Noninvasive detection and evaluation of carotid occlusive disease. *Arch. Surg.* 106:528–35
9. Barnes, R. W., Rittgers, S. E., Thornhill, B., Nix, L., Marszalek, P., Putney, W. 1979. Noninvasive cerebrovascular screening techniques: Indirect versus direct carotid interrogation. In *Smoking and Arterial Disease,* ed. R. Greenhalgh. In press
10. Gee, W., Oller, D. W., Wylie, E. J. 1976. Noninvasive diagnosis of carotid occlusion by ocular pneumoplethysmography. *Stroke* 7:18–21
11. Gee, W., Mehigan, J. T., Wylie, E. J. 1975. Measurement of collateral cerebral hemispheric blood pressure by ocular pneumoplethysmography. *Am. J. Surg.* 130:120–27
12. Blackshear, W. M., Thiele, B. L., Harley, J. D., Chikos, P. M., Strandness, D. E. 1979. A prospective evaluation of oculoplethysmography and carotid phonoangiography. *Surg. Gynecol. Obstet.* 148:201–5
13. Kobayashi, S., Hollenhorst, R. W., Sundt, T. M. Jr. 1971. Retinal arterial pressure before and after surgery for carotid artery stenosis. *Stroke* 2:569–75
14. Cohen, D. M., Wangelin, R., Trotta, C., Beven, E. G., Humphries, A. W., Young, J. R. 1975. Carotid compression tonography. *Stroke* 6:257–62
15. Capistrant, T. D., Gumnit, R. J. 1973. Detecting carotid occlusive disease by thermography. *Stroke* 4:57–64
16. Duncan, G. W., Gruber, J. O., Dewey, C. F. Jr., Myers, G. S., Lees, R. S. 1975. Evaluation of carotid stenosis by phonoangiography. *N. Engl. J. Med.* 293: 1124–28
17. Rosen, R. M., Parthasarathy, S. P., Turner, A. F., Blankenhorn, D. H., Roschke, E. J. 1977. Phonoangiography by autocorrelation. *Circulation* 55: 626–33
18. Barnes, R. W., Rittgers, S. E., Thornhill, B., Nix, L., Putney, W. 1979. Noninvasive determination of carotid artery operability by Doppler ultrasound. *Va. Med. Mon.* In press
19. Rutherford, R. B., Hiatt, W. R., Kreutzer, E. W. 1977. The use of velocity wave form analysis in the diagnosis of carotid artery occlusive disease. *Surgery* 82:695–702
20. Reinertson, J. E., Barnes, R. W. 1976. Carotid flow velocity scattering: diagnostic value in Doppler ultrasound arteriography. *Clin. Res.* 24(4):594A
21. Hokanson, D. E., Mozersky, D. J., Sumner, D. S., Strandners, D. E. Jr. 1971. Ultrasonic arteriography: a new approach to arterial visualization. *Biomed. Eng.* 6:420
22. Reid, J. M., Spencer, M. P. 1972. Ultrasonic Doppler technique for imaging blood vessels. *Science* 176:1235–36
23. Barnes, R. W., Bone, G. E., Reinertson, J., Slaymaker, E. E., Hokanson, D. E., Strandness, D. E. 1976. Noninvasive ultrasonic carotid angiography: Prospective validation by contrast arteriography. *Surgery* 80:328–35
24. Green, P. S., Marich, K. W. 1975. Realtime orthographic ultrasonic imaging for cardiovascular diagnosis. In Harrison D. E., Sandler, H., Miller, H. A. 1975. Cardiovascular imaging and image processing: Theory and practice. Palos Verdes Estates, Calif: Soc. Photoopt. Instrum. Eng. 72:31–35
25. Barber, F. E., Baker, D. W., Nation, A. W. C., Strandness, D. E. Jr., Reid, J. M. 1974. Ultrasonic duplex echo-Doppler

scanner. *IEEE Trans. Biomed. Eng.* 21:109–13

26. Cowan, R. J., Maynard, C. D., Meschan, I., Janeway, R., Shigeno, K. 1973. Value of the routine use of the cerebral dynamic radioisotope study. *Nucl. Med.* 107:111–16
27. Bone, G. E., Barnes, R. W. 1976. Clinical implications of the Doppler cerebrovascular examination: A correlation with angiography. *Stroke* 3:271–74
28. Garrett, W. V., Slaymaker, E. E., Barnes, R. W. 1979. Noninvasive perioperative monitoring of carotid endarterectomy. *J. Surg. Res.* 26:255–61
29. Ford, J. J., Baker, W. H., Ehrenheft, J. L. 1975. Carotid endarterectomy for non-hemispheric transient ischemic attacks. *Arch. Surg.* 110:1314–17
30. Baker, W. H., Barnes, R. W. 1977. The cerebrovascular Doppler examination in patients with non-hemispheric symptoms. *Ann. Surg.* 186:190–92
31. Javid, H., Ostermiller, W. E., Hengesh, J. W., Dye, W. S., Hunter, J. A., Najafi, H., Julian, O. C. 1971. Carotid endarterectomy for asymptomatic patients. *Arch. Surg.* 102:389–91
32. Thompson, J. E., Patman, R. D., Talkington, C. M. 1978. Long-term outcome of patients having endarterectomy compared with unoperated controls. *Ann. Surg.* 188:308–16
33. Cooperman, M., Martin, E. W. Jr., Evans, W. E. 1978. Significance of asymptomatic carotid bruits. *Arch. Surg.* 113:1339–40
34. Moore, W. S., Boren, C., Malone, J. M., Roon, A. J., Eisenberg, R., Goldstone, J., Mani, R. 1978. Natural history of nonstenotic, asymptomatic ulcerative lesions of the carotid artery. *Arch. Surg.* 113:1352–59
35. Kartchner, M. M., McRae, L. P. 1977. Noninvasive evaluation and management of the "asymptomatic" carotid bruit. *Surgery* 82:840–47
36. Treiman, R. L., Foran, R. F., Shore, E. H., Levin, P. M. 1973. Carotid bruit. *Arch. Surg.* 106:803–5
37. Lefrak, E. A., Guinn, G. A. 1974. Prophylactic carotid artery surgery in patients requiring a second operation. *South. Med. J.* 67:185–89
38. Carney, W. I. Jr., Stewart, W. B., DePinto, D. J., Mucha, S. J., Roberts, B. 1977. Carotid bruit as a risk factor in aortoiliac reconstruction. *Surgery* 81:567–70
39. Evans, W. E., Cooperman, M. 1978. The significance of asymptomatic unilateral carotid bruits in preoperative patients. *Surgery* 83:521–22
40. Bernhard, V. M., Johnson, W. D., Peterson, J. J. 1972. Carotid artery stenosis. *Arch. Surg.* 105:837–40
41. Hertzer, N. R., Loop, F. D., Taylor, P. C., Beven, E. G. 1978. Staged and combined surgical approach to simultaneous carotid and coronary vascular disease. *Surgery* 84:803–11
42. Ennix, C. L., Lawrie, G. M., Morris, G. C. Jr., Crawford, E. S., Howell, J. F., Reardon, M. J., Weatherford, S. C. 1979. Improved results of carotid endarterectomy in patients with symptomatic coronary disease: An analysis of 1,546 consecutive carotid operations. *Stroke* 10:122–25
43. Brewster, D. C., Abbott, W. M., Darling, R. C., Reidy, N. C., Raines, J. K. 1978. Cardiovascular surgery. *Circulation* 58:Suppl. 1, pp. I5–I9
44. Levin, S. M., Sondheimer, F. K. 1973. Stenosis of the contralateral asymptomatic carotid artery-to operate or not. *Vasc. Surg.* 7:3–13
45. Humphries, A. W., Young, J. R., Santilli, P. H., Beven, E. G., deWolfe, V. G. 1976. Unoperated, asymptomatic significant internal carotid artery stenosis: A review of 182 instances. *Surgery* 80:695–98
46. Johnson, N., Burnham, S. J., Flanigan, D. P., Goodreau, J. J., Yao, J. S. T., Bergan, J. J. 1978. Carotid endarterectomy: A follow-up study of the contralateral non-operated carotid artery. *Ann. Surg.* 188:748–52

Ann. Rev. Med. 1980. 31:219–32
Copyright © 1980 by Annual Reviews Inc. All rights reserved

LEGIONNAIRES' DISEASE

❖7355

Richard D. Meyer, M.D. and Sydney M. Finegold, M.D.[1]

Infectious Disease Section, Veterans Administration Wadsworth Medical
Center, Los Angeles, California 90073; and Department of Medicine,
School of Medicine, University of California, Los Angeles, California 90024

INTRODUCTION

A dramatic outbreak of serious pulmonary infection in Philadelphia in July, 1976, at a convention of the American Legion led to recognition of a discrete entity, Legionnaires' disease (L.D.), and eventually to isolation and identification of the etiologic agent (1–3). Stored sera from previously obscure outbreaks were tested by indirect fluorescent antibody technique, and *Legionella pneumophila* was documented as the cause of several other major outbreaks of disease since 1965 (2, 4, 5). The disease may occur sporadically or as part of an outbreak. Person-to-person transmission has not been documented.

The development and application of techniques of direct identification and culture (6, 7) have increased our understanding of the disease. The spectrum of the disease extends from asymptomatic seroconversion through a mild self-limited febrile illness without pneumonia (Pontiac Fever) (4) to the severe progressive pneumonia with significant mortality known as L.D. (1). The incubation period is usually 2–10 days but may be as long as 28 days (1).

By November 7, 1978, the Center for Disease Control (CDC) had confirmed 496 sporadic cases and 558 cases associated with outbreaks of L.D. in the United States. Cases have occurred in 43 states and the District of Columbia, and in Canada, Scotland, England, Wales, the Netherlands, Sweden, Denmark, Spain, Italy, Israel, and Australia. The prevalence of L.D. as the cause of sporadic, otherwise undiagnosed nosocomial pneumonias appears to be 3.8–10% (8, 9).

---

[1] Supported by the Medical Research Service of the Veterans Administration.

0066-4219/80/0401-0219$01.00

ETIOLOGY AND PATHOGENESIS

Legionella pneumophila is a Gram-negative weakly oxidase-positive, catalase-positive, fastidious bacterium. The optimum growth temperature is 35°C and pH 6.9–7.0. *Legionella* grows best in an atmosphere of 2.5–5% carbon dioxide, although it will grow in ambient air. It does not grow anaerobically (6).

The organism has not been grown on any medium that does not contain cysteine. It can be grown *in vitro* only with specially enriched culture media. The best medium designed to date is charcoal yeast extract agar. On artificial culture media, growth may not occur for up to 10 days. A brown soluble pigment (a melanin) is produced. The organism can be isolated by intraperitoneal inoculation of guinea pigs and by inoculating the yolk sac of embryonated hen's eggs (2). *Legionella* does not reduce nitrates, utilize carbohydrates, or attack urea. It does not appear to possess decarboxylases for either lysine or ornithine nor an arginine dihydrolase. A beta-lactamase is produced.

Microscopically, *L. pneumophila* rods are usually 2–3 μm long, often with pointed ends. Although the organism is Gram-negative, it does not take the ordinary Gram stain well. It does stain well with the Giménez stain and in tissue with silver impregnation stains (3).

Gas chromatography reveals large amounts of branched chain fatty acids in the bacterial cells. *L. pneumophila* is genetically unrelated at both the species and genus levels to all organisms tested with DNA hybridization.

There appear to be no strain differences to account for the variations in clinical characteristics among the different outbreaks. However, the organisms do fall into four serologic groups based on antigenic differences. The majority of isolates, including those from the Philadelphia outbreak of 1976, belong to Group I (prototype strain, Knoxville), with fewer isolates in Group II (prototype strain, Togus), Group III (prototype strain, Bloomington), and Group IV (prototype strain, Los Angeles) (10).

The pathogenesis of the disease remains little understood. The renal failure that may be observed is at least in part due to such factors as myoglobinuria, shock, and other prerenal factors; the encephalopathy is at least partly related to hypoxia and uremia. However, these factors clearly do not account for all of the renal and central nervous system manifestations of the disease. It has been speculated that there is generalized sepsis or release of a toxin from the organism. Our group has isolated *L. pneumophila* on *in vitro* culture of the blood from three patients with L.D. The organism contains an endotoxin but it is not highly active biologically. An exotoxin has also been demonstrated, but its significance is unknown.

PATHOLOGY

Pathologic findings are generally limited to the lungs (1, 3, 11). Varying degrees of lobar consolidation, fibrinous pleuritis, and serous or serosanguineous pleural effusions are common. Microscopic findings are acute fibrinopurulent bronchopneumonia around respiratory bronchioles. The alveolar septa are usually intact. Polymorphonuclear leukocytes predominate in an extensive dense eosinophilic intraalveolar proteinaceous and fibrinous exudate (11). Macrophages are frequently found. Acute diffuse alveolar damage with hyaline membrane formation and regenerating type 2 pneumocytes may be due to the disease or to oxygen toxicity.

The usual tissue stains do not show the bacilli. These include the Brown-Brenn, Gomori methenanine silver stain, Gridley fungus, and periodic acid-Schiff stains (3, 11). The Giménez stain is effective with fresh or fixed lung (3). The Dieterle silver impregnation stain (3) and Warthin-Starry stain show *L. pneumophila* well but are not specific. Organisms are seen in the exudate and inside polymorphonuclear leukocytes and alveolar macrophages and, on electron microscopic examination, inside phagocytic vacuoles. Specific diagnosis should be made by culture or with direct fluorescent antibody test (10). Biopsies done during the first week of illness show acute bronchopneumonia and interstitial pneumonia before oxygen therapy (12, 13).

EPIDEMIOLOGY

L. pneumophila is likely a soil and water organism transmitted by the air. Epidemiologic investigations in 1976 in Philadelphia indicated common-source airborne exposure. The attack rate was 9.0% for delegates in the hotel (1). *L. pneumophila* was isolated from lung tissue of four cases with *in vivo* techniques (2). Then five other Philadelphia patients who had not attended the convention were found to have had L.D. (14).

The earliest known sporadic case of L.D. occurred with bacteremia in 1947; the earliest known outbreak occurred in the summer of 1965 at St. Elizabeth's Hospital in Washington, D.C., with 81 ill patients; 57 had pneumonia and 14 died. The attack rate was highest for males, those with grounds privileges, or with beds next to open windows or near excavation sites (5).

In 1968, the explosive outbreak of Pontiac fever affected 144 people in a new county health department building (4). The grounds recently had been graded. The air-conditioner had large leaks from the evaporative condensor discharge units, which drained into the air ducts. When the building reopened after repairs, no further cases occurred. Available paired serum specimens showed seroconversion. *L. pneumophila* Serogroup I was recovered from lung nodules of immunosuppressed guinea pigs exposed to air in the building and to aerosols from the evaporative condensor (4, 10).

Tourists from Scotland and other parts of the United Kingdom have accounted for at least 10 cases with five deaths after staying at the same hotel in Benidorm, Spain, between 1973 and 1978. Other tourists from western and northern Europe have acquired Legionnaires' disease on Mediterranean cruises and trips to Corfu, France, Germany, Italy, and Portugal (13).

Retrospective analysis of a different convention held in 1974 at the same Philadelphia hotel involved in 1976 showed that L.D. occurred in 2.9% of 392 respondents. This was lower than the 4.0% overall attack in 1976 (1, 15). In Nottingham, England, 19 cases with two deaths from 1976 to 1978 have been identified. Nottingham and the hotels previously cited may be endemic areas.

In Burlington, Vermont, 69 confirmed and presumptive cases of L.D. with 17 deaths were found in 1977 (12). At least 13 were hospital-acquired cases but the majority were community-acquired. Another nine cases with underlying chronic disease or immunosuppression were found in the summer of 1978 and accounted for 11.6% of nosocomial pneumonias (16). These data and the 26% seroreactivity rate in the community suggest that Burlington is another endemic area (12).

The new Veterans Administration (VA) Wadsworth Medical Center opened in March 1977 and intermittent excavation followed. From May 1977 to August 1979, 111 hospital-acquired cases with 17 deaths due to the infection itself were confirmed (9, 13, 17). The first case was identified in a renal transplant recipient. Renal transplantation has been temporarily suspended because 6 of 12 renal transplant recipients developed L.D. in 1977 (9). Cases cluster in the late fall and early winter but occur throughout the year. All have occurred in inpatients, clinic patients, or employees during the known incubation period (9, 13). The median interval from admission to onset for inpatients was 17 days. Duration of hospitalization was a risk factor except for immunosuppressed patients who had a nine-fold greater risk of contracting L.D. Immunosuppression was a significant risk factor ($P < 0.01$). Eight cases (0.5% attack rate) occurred in a cohort of 1658 consecutive admissions followed over three months beginning in February 1978. Eight others who seroconverted were asymptomatic. Legionnaires' disease accounted for 8 of 73 cases of nosocomial pneumonia and for 4 of 14 fatal nosocomial pneumonias (9). No other clusters have been identified in the area.

Prevalence of a positive indirect fluorescent antibody titer ⩾ 1 : 128 was significantly more common in groundskeepers and in employees who work inside the building compared to employees at an office building less than 0.5 km away. Environmental isolation from the cooling tower water was accomplished by the CDC. Cases continue to occur despite effective chlorination and negative cultures of this water.

Nine cases of Legionnaires' disease were identified in Columbus, Ohio, in 1977, associated with exposure to two hospitals with recent excavation. In Kingsport, Tennessee, in August and September, 1977, 33 cases were identified in a hospital and in the community. Then new cases ceased.

In Bloomington, Indiana, 39 cases have been identified, with four deaths; 35 patients had stayed at least one night between May 1977 and August 1978 in a hotel in the Student Center on campus (18). The estimated attack rate for overnight guests appeared to be 0.12%. Seropositivity in employees indicated that the hotel is an endemic site. Environmental isolation of three serogroups has been achieved from mud, creek water, and cooling tower water. Cases have occurred in the winter when the cooling tower was drained and not operating (18).

A cluster of eight cases in golfers at an Atlanta, Georgia, country club in 1978 where *L. pneumophila* was cultured from water from an evaporative condensor on the golf course and from autopsy lung is under investigation, as is a cluster of up to 33 confirmed and presumptive cases in a Memphis hospital and the nearby area. In Memphis *L. pneumophila* in an auxiliary water tower probably was blown to the street and hospital.

An outbreak of L.D. with approximately 38 confirmed and presumptive cases was identified in the garment district of New York City in August and September 1978. Seropositivity in control populations (29% ⩾ 1 : 128) in the area indicates that it is an endemic area. An outbreak from 1976 through 1978 in Norwalk, Connecticut, and two summer 1979 outbreaks in Eau Claire, Wisconsin, and Jamestown, New York, are being investigated by the CDC.

The sporadic cases confirmed by the CDC extend over most of the U.S. Most have been in the Northeast and Middlewest with a late summer to early fall peak incidence. The age of the patients has ranged from 2 to 84 years with a median of 54; males predominate. Sporadic cases have been recognized in Australia, Canada, England, Wales, the Netherlands, Israel, Italy, Scotland, Denmark, Sweden, and Norway, as well as in tourists (13). A possible cluster of sporadic cases where some have been nosocomial is under study in Detroit (19). A CDC study of nosocomial pneumonia in 40 hospitals showed that L.D. accounted for 3.8% of fatal cases (8). This suggests that 950 cases of fatal nosocomial L.D. pneumonia occur yearly in the United States. Corticosteroid administration, temperature exceeding 39.2°C, and occurrence between June and November were common findings (8). Since the Bloomington outbreak included at least five cases in people from other states thought to have been sporadic, other sporadic cases may also be linked to outbreaks.

PREDISPOSING FACTORS

Geographic factors, reflected by outbreaks and by rates of seropositivity, interact with host factors in predisposition to acquisition and the severity of the disease. Late middle-aged to elderly males, naturally, predominated in the Philadelphia (1) and Los Angeles outbreaks (9, 13, 17). Males also predominated in other outbreaks (5, 12, 13, 18), in sporadic nosocomial cases (8), and in other sporadic cases (20).

The greater risk of acquisition of L.D. in immunocompromised patients in various hospitals is noteworthy (8, 12, 13, 17, 19). Chronic dialysis patients have contracted Legionnaires' disease in diverse hospitals (13). Of 241 reported cases, 63% had significant underlying disease. Underlying diseases in 47 immunosuppressed patients were neoplasms, renal transplantation, and, less frequently, rheumatologic diseases. Other common underlying diseases were organic heart disease, lung disease, renal disease, and diabetes mellitus. Alcohol abuse and smoking were common (13) and were risk factors in sporadic cases (20). An excavation site near a home, an overnight trip within the last two weeks (20), and possibly the use of a home air-conditioner are risk factors (12).

CLINICAL FEATURES

Most recognized cases of L.D. have had pneumonia, usually severe. However, the disease has a broad spectrum. Asymptomatic infection has been confirmed by diagnostic seroconversion (9). An outbreak at the Pontiac, Michigan Health Department in 1968 involved a relatively mild disease without pneumonia (4). In the 1976 Philadelphia outbreak, 13 patients who fulfilled epidemiologic criteria for L.D. had normal chest radiographs (1). It is likely that reported cases of L.D. underestimate the prevalence of the disease and do not accurately portray its clinical spectrum.

Pneumonia

The onset of the disease is usually not abrupt. Initial symptoms often include malaise, weakness, lethargy, headache, and anorexia and sometimes myalgia. Upper respiratory symptoms are usually absent (1, 13, 17). Cough is present in about 90% of patients. It is initially dry in most patients, but may become productive, usually of nonpurulent or minimally purulent secretions, in half the cases. Hemoptysis, usually mild, is not uncommon. Three quarters of patients manifest shaking chills, often recurrently (13, 17).

In our experience at the VA Wadsworth Medical Center (13), watery diarrhea was seen in almost half the cases. Nausea and vomiting are seen in about one fourth of patients. Headache is seen in one fourth to one half of patients. Sweats, sometimes profuse and drenching, are seen in one fourth of patients. Myalgia and/or arthralgia are reported in about half of the patients. One third of patients have pleurisy.

The physical findings in pneumonic Legionnaires' disease are usually impressive. Most patients have temperatures greater than 39°C and half to two thirds of patients have temperatures greater than 40°C, even when receiving corticosteroids or antipyretic agents. The fever tends to be nonremitting. Relative bradycardia is seen in almost two thirds of patients.

Early, only rales may be noted but the classic case goes on to frank consolidation. Evidence of a small pleural effusion or a pleural rub may be noted. Striking changes in mental status of the patient are noted in from one fourth to one third of individuals. These include confusion, disorientation, lethargy, depression, emotional lability, insomnia, hallucinations, delirium, and retrograde amnesia (13). Focal neurological findings have been noted rarely.

The disease extends to involve adjacent or other lobes of the lung, particularly if appropriate therapy is not started. Cavitation or abscess formation has been reported in a few cases (11, 13, 16). Additional complications are pneumatocele, pneumothorax, empyema, and adult respiratory distress syndrome. L.D. may coexist with other pulmonary infection (13) in which case the clinical picture may be much less distinctive.

Extrapulmonary Manifestations

The occurrence of certain central nervous system signs has been noted earlier. Additional central nervous system manifestations that have been reported include slurred speech, coma, memory defects, fine or coarse tremors, stiff neck, ataxia, cerebellar dysfunction, and peripheral neuropathy. Two cases of L.D. with associated severe cerebral disturbance have been reported (21). Severe aphasia and ataxia persisted for weeks in one case and were still present to a minor degree after three years. The second case had persistent aphasia after five months. Ataxia persisted for at least three years in another case.

The liver is at times enlarged and tender and clinical jaundice is seen occasionally. On occasion, guaiac-positive stools or frank gastrointestinal bleeding has been noted.

Abnormalities related to the kidneys range from proteinuria and hematuria to overt renal failure (1, 15). Hyaline and granular casts have been seen periodically and red blood cell casts rarely. Renal failure related to shock, myoglobinuria, or acute tubulointerstitial nephritis has been reported. Dialysis has been required at times. One patient had mild residual renal insufficiency for at least nine months.

Elevation of serum creatine kinase and aldolase in a few patients has been seen in relation to rhabdomyolysis or myositis. Hypotension or shock has been noted occasionally. Hematologic abnormalities include disseminated intravascular coagulation, thrombocytopenia, and severe anemia. Inappropriate secretion of antidiuretic hormone has been noted (17).

Pontiac Fever

The Pontiac Fever outbreak affected 95% of employees. It was a relatively mild, self-limited illness consisting primarily of chills, low-grade fever, headache, and myalgia lasting two to five days.

The mean incubation period was only 36 hours. The entire symptom complex appeared within six to twelve hours of onset. Other symptoms commonly noted were dizziness, slight cough, photophobia, nausea, and diarrhea. Sore or dry throat was present in one third of patients and midsternal chest pain or constrictive sensation was noted by 45%. Also noted were confusion, poor coordination, nightmares, insomnia, difficulty with memory or concentration, and general irritability. Chest radiographs, performed in 17 individuals, were negative. No significant residual defects were noted and there were no deaths (4).

DIAGNOSIS

Distinctive Clinical Features

Although *L. pneumophila* may cause asymptomatic infection or mild illness, the classical presentation is that of pneumonia. In the presence of an outbreak, of course, all cases of pulmonary infection must be considered as possibly due to *L. pneumophila.* Predisposing factors, such as immunosuppressive therapy, should also raise the question of L.D. in an individual with pneumonia. With involvement of other organ systems such as the central nervous system, liver, gastrointestinal tract, and/or kidneys, L.D. becomes even more likely.

The finding of only small numbers of polymorphonuclear or mononuclear leukocytes and few or no organisms in sputum or transtracheal aspirate suggests L.D., as does the absence of pathogens on routine culture of such specimens, or pleural fluid, in an individual with pneumonia. High fever, particularly with relative bradycardia, is an important clue. Evidence of an encephalopathy and/or renal abnormalities and diarrhea similarly suggest L.D. Significant myalgia in a patient with pneumonia would be suggestive, as would hemoptysis with or without pleural pain. When otherwise unexplained abnormal liver function tests or hypophosphatemia are seen in conjunction with an acute pneumonia, L.D. must be strongly considered.

Failure of an apparently bacterial pneumonia to respond to agents such as penicillins, cephalosporins, or aminoglycosides should raise the possibility of L.D. Conversely, an impressive response to erythromycin suggests this diagnosis if other causes have been excluded.

Routine Laboratory Diagnosis

Acute pneumonia is the striking feature of almost all cases of L.D. The first radiographic finding usually is a patchy alveolar infiltrate in one lobe (1, 13, 15). Consolidation in the area of the initial infiltrate is common. Spread to other areas of the ipsilateral lung is noted next. Infiltrates remain unilateral in about half to three fourths of patients (13). In others there is contralateral spread. Consolidation and lower lobe involvement are common (1, 13, 17). Infiltrates may progress despite appropriate therapy and eventual clinical recovery. Patients who survive after therapy usually show radiographic improvement within two weeks but clearing may be delayed (13, 15). Pleural effusion may be found before the infiltrate; this occurred in about half of our patients (13, 17). An initial diffuse infiltrate is distinctly rare or unknown. Less common findings are nodular infiltrates and cavity formation (11, 13, 16, 19).

Leukocytosis (WBC $\geqslant$10,000/mm^3) is found in half to three fourths of patients (13). Immature forms may be noted (13, 15). Leukopenia, a bad prognostic sign, is uncommon (1, 15) as is thrombocytopenia (13). Disseminated intravascular coagulation has occurred several times. Increase in erythrocyte sedimentation rate, serum levels of lactic dehydrogenase, alanine aminoaspartase (SGOT), and alkaline phosphatase are common, and modest bilirubin elevation less common (1, 13, 15, 17).

Hematuria was found in 16% of reported cases but has not been found at Wadsworth (13). Proteinuria occurs in about half and hyaline or granular or red blood cell casts less frequently (13). Acute renal failure with myoglobinuria (14) and nonoliguric renal failure that took nine months to resolve have been noted.

Hyponatremia from the syndrome of inappropriate antidiuretic hormone occurs in about half (13). Hypophosphatemia was noted in half of our patients and reflects the accompanying Gram-negative septicemia (13, 17). Uncommon findings are elevations of creatine kinase and adolase (14, 19). Arterial blood gas abnormalities reflect the extent of pneumonic involvement. Spinal fluid examination is normal (17). Positive cold agglutinin and mycoplasma complement fixation tests have been found in a few patients with confirmed L.D. (13, 17).

Serum specimens for the indirect fluorescent antibody test should be obtained at onset of illness and preferably weekly for six weeks. The antigen currently routinely used is the Philadelphia or Serogroup I. About two thirds of patients show seroconversion at three weeks and all by six weeks if seroconversion is used for diagnosis (17). Current criteria for confirmation of diagnosis are a titer quadrupled or more to $\geqslant 1:128$ with compatible clinical disease. A convalescent titer of $\geqslant 1:256$ with compatible clinical disease is presumptive; yet in endemic areas, interpretation of a single titer may be difficult. Known cross-reactions occur with tularemia, plague, and leptospirosis. Use of culture of *L. pneumophila* and/or demonstration by direct immunofluorescence on respiratory tract secretions reveals that not all patients in our hospital show seroconversion on serial testing against all four serogroups.

Rapid diagnosis should be attempted with examination of expectorated sputum, transtracheal aspirates, endotracheal tube aspirates, bronchoscopic washings, lung biopsies, lung aspirates, or pleural fluid by the direct immunofluorescent test (7, 13, 16, 19, 22). This appears very useful.

L. pneumophila has been recovered from autopsy lung specimens (2, 13), transtracheal aspirates (22, unpublished), sputum (unpublished), pleural fluid (23), lung biopsies (13, 19), and in our hospital three times from blood using *in vitro* techniques. At Wadsworth, 25 clinical isolates, mostly Group I, have been recovered.

Gram stains of most transtracheal aspirates have shown only a few polymorphonuclear leukocytes and/or macrophages and less commonly faintly staining Gram-negative bacilli (13, 17). Exudative pleural effusions are more common than are transudates (17).

Special culture techniques in reference laboratories for clinical or environmental specimens include yolk sac inoculation of embryonated hen's eggs and intraperitoneal injection of guinea pigs (2).

DIFFERENTIAL DIAGNOSIS

Clinical diagnosis outside an outbreak situation may be difficult. Progressively rising fevers, persistent and recurring rigors, nonproductive cough, prostration, and relative bradycardia and extrapulmonary manifestations should suggest L.D. Liver function abnormalities and hypophosphatemia are clues also. An unimpressive cough and scant sputum in the face of considerable pneumonia is notable. Recurrent rigors, rather than single chill, favor L.D. over uncomplicated *Streptococcus pneumoniae* pneumonia. A persistent hacking cough is more common with *Mycoplasma pneu-*

moniae infection, which usually also occurs in younger people but there is some overlap of ages.

Immunosuppression, high fever, and corticosteroid use in nosocomial cases and late summer onset in some areas are helpful. The failure to isolate other bacterial pathogens is suggestive but aspiration pneumonia or other dual infections have been also noted (13, 16, 17).

One must also consider influenzae, adenovirus, psittacosis, and Q fever. The chest radiograph in psittacosis normally shows a smaller infiltrate than in L.D. and in Q fever the infiltrate is more likely to be a diffuse interstitial one. Tularemia or plague with pneumonia must also be considered in sporadic cases.

THERAPY

The usual supportive measures required for seriously ill patients are, of course, essential. Reduction in the dosage of immunosuppressive agents, particularly corticosteroids, in patients with L.D. should be considered (9).

L. pneumophila produces a beta-lactamase which is basically a cephalosporinase but which also affects penicillins. In *in vitro* studies and studies involving chick embryo cells, yolk sac culture, and intraperitoneally infected guinea pigs, several drugs showed good activity against *L. pneumophila.* Overall, rifampin and erythromycin showed the greatest activity (24).

Some of the guinea pigs exposed in the Pontiac fever outbreak of 1968 were treated prophylactically with tetracycline and streptomycin. These animals had a 38% reduction in incidence of pneumonia as compared to untreated exposed guinea pigs (25).

In the epidemic of L.D. in Philadelphia in 1976 (1), the case-fatality ratio was high in patients treated with cephalothin (20/49); intermediate in patients treated with aminoglycosides (9/25), chloramphenicol (3/10), ampicillin (10/41), and penicillin (6/30); and was most favorable in patients given tetracycline (3/30) or erythromycin (2/18).

The experience with tetracycline is confusing. In the Philadelphia outbreak, as noted, it appears that tetracycline was relatively effective (1, 15). However, several patients have developed the disease while receiving tetracycline or have had clearcut progression during tetracycline therapy (17). Penicillins and cephalosporins, alone or with aminoglycosides, have not provided a favorable effect in L.D. (1, 13, 15, 17, 19). A few reports of therapy with gentamicin plus either clindamycin or metronidazole indicated no favorable response. Chloramphenicol failed in one case.

Our initial experience (17) was that most cases of L.D. treated with drugs other than erythromycin progressed. Three patients received tetracycline; of these, two appeared to respond. Except for one patient who apparently responded to high dose ampicillin therapy, the following drugs were used without effectiveness: ampicillin, penicillin, carbenicillin, oxacillin, cephalothin, cephalexin, cefoxitin, cefazolin, amikacin, gentamicin, and clindamycin. Our larger experience (13) impresses us that erythromycin is an effective drug and generally leads to prompt clinical improvement (mortality 13% in patients treated with erythromycin vs 55% in patients not treated with this drug). We have used doses of two or four grams daily by either the oral or intravenous routes. We favor the intravenous route and the higher dosage for patients who are very ill with L.D. There is one report of failure of oral erythromycin therapy in three patients, one of whom developed the disease while receiving erythromycin (26). Patients should be treated for at least three weeks because relapse or protracted convalescence has been noted in patients treated for shorter period.

There are a few instances of successful use of rifampin together with tetracycline or erythromycin. Rifampin should not be used as a single agent because rapid one-step resistance may develop. Cotrimoxazole in high dosage may have been effective in two patients in our series.

PROGNOSIS

Case-fatality rates and prognosis are greatly affected by underlying disease and therapy administered. A few patients not treated appropriately have recovered after about six to eight days of severe illness (1, 15). Immunosuppressed patients at Wadsworth who received erythromycin had a case-fatality rate of 24% and those who did not 30%. In nonimmunosuppressed patients, the case-fatality rate was 7% with erythromycin therapy and 25% in those who did not receive erythromycin (13).

Patients who apparently recover commonly complain of easy fatigability, malaise, and weakness for weeks to months and may show persistent central nervous system abnormalities (13, 15, 21). Pulmonary function tests were unchanged in some not treated with erythromycin except for a decrease in diffusing capacity (27).

SUMMARY

Legionella pneumophila infections frequently manifest themselves as a multisystem disease with acute pneumonia. Certain clinical and laboratory features are helpful in diagnosis but none are pathognomonic. Diagnosis frequently must be made clinically and erythromycin given presumptively because of the delay in seroconversion but culture and direct fluorescent antibody testing are quite useful for rapid diagnosis. Epidemiologic and laboratory investigations will undoubtedly yield considerable information about this fascinating bacterial disease.

Literature Cited

1. Fraser, D. W., Tsai, T. F., Orenstein, W., Parkin, W. E., Beecham, H. J., Sharrar, R. G., Harris, J., Mallison, G. F., Martin, S. M., McDade, J. E., Shepard, C. C., Brachman, P. S., Field Investigation Team. 1977. Legionnaires' disease: description of an epidemic of pneumonia. *N. Engl. J. Med.* 297:1189–97
2. McDade, J. E., Shepard, C. C., Fraser, D. W., Tsai, T. F., Redus, M. A., Dowdle, W. R., Laboratory Investigation Team. 1977. Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease. *N. Engl. J. Med.* 297:1197–1203
3. Chandler, F. W., Hicklin, M. D., Blackmon, J. A. 1977. Demonstration of the agent of Legionnaires' disease in tissue. *N. Engl. J. Med.* 297:1218–20
4. Glick, T. H., Gregg, M. B., Berman, B., Mallison, G., Rhodes, W. W., Kassanoff, I. 1978. Pontiac fever—epidemic of unknown etiology in a health department. 1. Clincal and epidemiologic findings. *Am. J. Epidemiol.* 107:149–60
5. Thacker, S. B., Bennett, J. V., Tsai, T. F., Fraser, D. W., McDade, J. E., Shephard, C. C., Williams, K. H. Jr., Stuart, W. H., Dull, H. B., Eickhoff, T. F. 1978. An outbreak in 1965 of severe respiratory illness caused by the Legionnaires' disease bacterium. *J. Infect. Dis.* 138: 512–19
6. Jones, G. L., Hebert, G. A., eds. 1978. *Legionnaires': The Disease, the Bacterium and the Methodology.* Washington: Cent. for Dis. Control, Public Health Serv., US Dept. Health, Educat., Welfare. 195 pp.
7. Broome, C. V., Cherry, W. B., Winn, W. C. Jr., MacPherson, B. R. 1979. Rapid diagnosis of Legionnaires' disease by direct immunofluorescent staining. *Ann. Intern. Med.* 90:1–4
8. Cohen, M. L., Broome, C. V., Paris, A. L., Martin, W. T., Allen, J. R. 1979. Fatal nosocomial Legionnaires' disease: clinical and epidemiologic characteristics. *Ann. Intern. Med.* 90:611–13
9. Haley, C. E., Cohen, M. L., Halter, J., Meyer, R. D. 1979. Nosocomial Legionnaires' disease: a continuing common-source epidemic at Wadsworth Medical Center. *Ann. Intern. Med.* 90:583–87
10. McKinney, R. M., Thacker, L., Harris, P. P., Lewallen, K. R., Hebert, G. A., Edelstein, P. H., Thomason, B. M. 1979. Four serogroups of Legionnaires' disease bacteria by direct immunofluorescence. *Ann. Intern. Med.* 90: 621–24
11. Winn, W. C., Glavin, F. L., Perl, D. P., Keller, J. L., Andres, T. L., Brown, T. M., Coffin, C. M., Sensecqua, J. E., Roman, L. N., Craighead, J. E. 1978. The pathology of Legionnaires' disease: fourteen fatal cases from the 1977 outbreak in Vermont. *Arch. Pathol. Lab. Med.* 102:344–50
12. Broome, C. V., Goings, S. A. J., Thacker, S. B., Vogt, R. L., Beaty, H. N., Fraser, D. W., Field Investigation Team. 1979. The Vermont epidemic of Legionnaires' disease. *Ann. Intern. Med.* 90:573–76
13. Kirby, B. D., Snyder, K. M., Meyer, R. D., Finegold, S. M. 1979. Legionnaires' disease: report of sixty-five nosocomially acquired cases and review of the literature. *Medicine (Baltimore).* 59: In press
14. Friedman, H. M. 1978. Legionnaires' disease in non-legionnaires: A report of five cases. *Ann. Intern. Med.* 88:294–302
15. Tsai, T. F., Finn, D. R., Plikaytis, B. D., McCauley, W., Martin, S. M., Fraser, D. W. 1979. Legionnaires' disease: clinical features of an epidemic in Philadelphia. *Ann. Intern. Med.* 90:509–17
16. Gump, D. W., Frank, R. O., Foster, R. S., Winn, W. C., Broome, C. V., Cherry, W. B. 1979. Legionnaires' disease in patients with associated serious illness. *Ann. Intern. Med.* 90:538–42
17. Kirby, B. D., Snyder, K. M., Meyer, R. D., Finegold, S. M. 1978. Legionnaires' disease: Clinical features of 24 cases. *Ann. Intern. Med.* 89:297–309
18. Politi, B. D., Fraser, D. W., Mallison, G. F., Mohatt, J. V., Morris, G. K., Patton, C. M., Feeley, J. C., Telle, R. D., Bennett, J. V. 1979. A major focus of Legionnaires' disease in Bloomington, Indiana. *Ann. Intern. Med.* 90: 587–91
19. Saravolatz, L. D., Burch, K. H., Fisher, E., Madhavan, T., Kiani, D., Neblett, T., Quinn, E. 1979. The compromised host and Legionnaires' disease. *Ann. Intern. Med.* 90:533–37
20. Storch, G., Baine, W. B., Fraser, D. W., Broome, C. V., Clegg, H. W., Cohen, M. L., Goings, S. A. J., Politi, B. D., Terranova, W. A., Tsai, T. F., Plikaytis, B. D., Shepard, C. C., Bennett, J. V. 1979. Sporadic community-acquired Legionnaires' disease in the United States: a case-control study. *Ann. Intern. Med.* 90:596–600

21. Lees, A. W., Tyrrell, W. F., Boyd, J. F. 1977. Legionnaires' disease. *Lancet* 2:1187 (Letter)
22. Lattimer, G. L., McCrone, C., Galgon, J. 1978. Diagnosis of Legionnaires' disease from transtracheal aspirate by direct fluorescent antibody staining and isolation of the bacterium. *N. Engl. J. Med.* 299:1172–73
23. Dumoff, M. 1979. Direct *in vitro* isolation of the Legionnaires' disease bacterium in two fatal cases: cultural and staining characteristics. *Ann. Intern. Med.* 90:694–96
24. Lewis, V. J., Thacker, W. L., Shepard, C. C., McDade, J. E. 1978. In vivo susceptibility of the Legionnaires' disease bacterium to ten antimicrobial agents. *Antimicrob. Agents Chemother.* 13:419–22
25. Fraser, D. W., Wachsmuth, I. K., Bopp, C., Feeley, J. C., Tsai, T. F. 1978. Antibiotic treatment of guinea-pigs infected with agent of Legionnaires' disease. *Lancet* 1:175–78
26. Venkatachalam, K. K., Saravolatz, L. D., Christopher, K. L. 1979. Legionnaires' disease. A cause of lung abscess. *J. Am. Med. Assoc.* 241:597–98
27. Lattimer, G. L., Rhodes, L. V. III, Salventi, J. S., Galgon, J. P., Stonebraker, W., Boley, S., Haas, G. 1979. The Philadelphia epidemic of Legionnaires' disease: Clinical, pulmonary, and serologic findings two years later. *Ann. Intern. Med.* 90:522–26

Ann. Rev. Med. 1980. 31:233–38
Copyright © 1980 by Annual Reviews Inc. All rights reserved

PULMONARY FUNCTION DURING AND AFTER COMMON RESPIRATORY INFECTIONS

❖7356

William J. Hall, M.D. and R. Gordon Douglas, Jr., M.D.

Departments of Medicine and Pediatrics, University of Rochester
School of Medicine, Rochester, New York 14642

INTRODUCTION

Viral respiratory illnesses in adults account for over 50% of work and school absenteeism yearly in the United States. Recent work on the pathogenetic mechanisms of these infections has delineated a complex interaction between viral respiratory infection and lung function that may help explain the often protracted nature of the symptoms. Additionally, studies of pulmonary function during and after viral infection have defined some of the important defense mechanisms of the lung.

COMMON COLD

Recent evidence indicates that even "uncomplicated" viral upper respiratory infection in adults, i.e. infection in which clinical manifestations remain limited to the upper respiratory tract and in which chest roentgenograms and conventional pulmonary function tests are normal, are commonly associated with prolonged physiologic abnormalities suggestive of lower respiratory tract involvement. These include abnormal frequency dependence of dynamic compliance (1), diminished steady-state diffusing capacity (2), abnormal closing volumes (3), and diminished density-dependent expiratory flow rates (3). A prospective study (4) of young children with uncomplicated upper respiratory infection showed diminished forced expiratory flow rates—a particularly significant observation because other complicating factors such as long-term exposure to air pollutants and cigarette smoking were obviated.

0066-4219/80/0401-0233$01.00

These studies suggest that a common site of host-virus interaction is in the peripheral airways of man. Much recent attention has focused on physiologic abnormalities in this segment of the tracheobronchial tree. These airways, usually arbitrarily defined as less than 2 mm in diameter, normally contribute very little resistance to airflow in the lung, compared to the larger airways. Consequently, considerable narrowing of peripheral airways can be present without marked symptoms or abnormalities of conventional pulmonary function tests. For this reason, these airways have been dubbed the "silent zone" of the lung (5). Another consistent observation is the protracted duration of physiologic abnormalities, usually averaging 3–12 weeks, well beyond the duration of upper respiratory tract symptoms.

INFLUENZA INFECTION

Earlier studies evaluating physiologic impairment often included subjects with acute respiratory illness not precisely characterized according to specific viral etiology. Our approach has been to combine specific viral diagnosis with physiologic testing. For several years we have studied pulmonary function in otherwise healthy young adults with nonpneumonitic, naturally acquired influenza A virus infection. In our initial studies 13 subjects with proven influenza A/Port Chalmers/73 H_3N_2 infection had abnormal frequency dependence of total pulmonary resistance, which suggests a generalized increase in the resistance to air flow in the peripheral airways (6). As was noted with other viral infections, these abnormalities persisted beyond the period of symptomatic illness. In a subsequent epidemic of influenza A/Port Chalmers/73 H_3N_2 virus, we further documented peripheral airway dysfunction in uncomplicated influenza by demonstrating diminished density-dependent forced flow rates in 15 volunteers with naturally acquired disease (7). Similar findings were observed in 36 of 39 (92%) subjects with influenza A/Victoria/3/75 H_3N_2 infection (8).

These observed changes in pulmonary function brought on by influenza infection allow us to conveniently and objectively assess the efficacy of antiviral chemotherapy. The prophylactic effect of amantadine against influenza A virus belonging to the H_2N_2, H_3N_2, and H_1N_1 groups is well documented (9). In addition, signs and symptoms of influenza resolve more rapidly when amantadine is used therapeutically (9). In two randomized double-blind studies in our laboratory, patients with peripheral airway dysfunction improved more rapidly when treated with oral amantadine than did the untreated controls (7, 8). This correlates with more rapid improvement in signs and symptoms, more rapid defervescence, and more rapid decrease in virus shedding among amantadine-treated subjects.

Recent attention has focused on development of live attenuated influenza A virus vaccines. In light of the known physiologic changes associated with naturally acquired infection, evaluation of these intranasal-administered products should include physiologic testing. Although some investigators (10, 11) describe minor transient pulmonary function abnormalities of questionable clinical significance, others have been unable to confirm these findings (12, 13). Consequently, there is no convincing evidence that live influenza virus vaccines have serious deleterious effects on lung function. In tests of over 100 subjects who have received either temperature-sensitive, cold-adapted, or inhibitor-resistant viruses, we have not observed the peripheral airways abnormalities associated with naturally occurring influenza (unpublished observations).

In another study in chronic bronchitis, symptoms were more frequent after live vaccine compared to placebo, but other differences between the groups may have explained the observed symptoms (14). We did not observe differences in clinical responses among elderly and chronically ill subjects who received live attenuated virus compared to those who received killed vaccine (15).

PATHOPHYSIOLOGY OF ALTERED LUNG FUNCTION

The most characteristic histologic hallmark of viral respiratory infection is bronchial epithelial damage (16). While an inflammatory response to viral replication undoubtedly contributes substantially to the observed abnormalities in pulmonary function following viral infection, it seems unlikely that this is the sole mechanism. In our studies on nonpneumonic influenza, pulmonary function abnormalities are most pronounced 7–10 days after the onset of illness, usually at the time when clinical manifestations are improved. Also, abnormalities are prolonged beyond the usual period of viral shedding (6–8), which suggests an additional pathophysiologic mechanism along with inflammation. Recent evidence suggests that exaggerated bronchial muscle reactivity may be this additional factor. Empey and colleagues (17) demonstrated transient bronchial hyperreactivity to histamine aerosol and an exaggerated cough reflex to inhaled citric acid during and after uncomplicated influenza A infection. Both responses could be blocked by prior administration of atropine. These observations suggested that airway epithelial damage by infection results in a vagally mediated reflex causing generalized bronchospasm. It was postulated that the afferent receptors were in fact rapidly adapting airway epithelial receptors "sensitized" by epithelial damage (18). These receptors and associated pathways have been implicated in the hyperreactivity of asthmatics to a variety of stimuli, including histamine and citric acid.

We have investigated airway reactivity during specific respiratory viral infections by utilizing carbecholamine (Carbachol) as a cholinergic aerosol challenge. Subjects with nonpneumonitic influenza A infection manifest transient bronchial hyperreactivity for an average of three weeks following onset of symptoms (8). Furthermore, no relationship with an atopic predisposition could be discerned. Recently, we studied a naturally occurring outbreak of respiratory syncytial virus infection in a group of young, nonatopic adults (19). The major abnormality noted was exaggerated bronchial reactivity to cholinergic stimulation, in some cases persisting for over eight weeks.

There is preliminary evidence suggesting that respiratory viral infection may also alter the pulmonary response to environmental stimuli. We recently studied a group of nonatopic subjects with documented upper respiratory viral infection. Subjects demonstrated enhanced bronchoconstriction to exercise in cold air. Altered reactivity to this common environmental stimulus persisted for three weeks following mild infection (20). This observation suggests that viral infection may alter the response of the respiratory tract to incompletely conditioned air, and may explain the protracted duration of symptoms commonly noted with these infections.

Prior epidemiologic studies have noted a correlation between acute respiratory illness and elevated levels of air pollutants (21). Utell et al (22) recently demonstrated that particulate nitrate, a common air pollutant, produced no effect either in normals or asthmatics after short-term inhalation (22). However, when subjects with uncomplicated influenza infection were exposed to an identical nitrate concentration they uniformly demonstrated transient diminution in forced expiratory flow rates, lasting up to three weeks following infection (23). If these studies can be confirmed, the adverse health effects of automotive and industrial pollutants may have to be reevaluated in terms of their potential synergy with viral illness.

Other related studies suggest that viral respiratory infection may alter the balance between the autonomic nervous system and humoral agents, which together modulate airway muscular tone. Szentivani (24) demonstrated such autonomic imbalance, chiefly characterized by diminished beta-adrenergic response of bronchial smooth muscle. Busse (25) studied the beta-adrenergic sensitivity of asthmatic patients with intercurrent respiratory viral infection. The common beta-adrenergic compound, isoproterenol, normally inhibits lysosomal enzyme release in granulocytes, and furthermore, the degree of inhibition is markedly diminished in asthmatic patients compared to normal subjects. These investigators found an even further transient diminution of isoproterenol inhibition during intercurrent viral infection. They suggested that a similar change in beta-adrenergic tone of airways could partially explain exaggerated bronchoconstrictor responses observed during viral infections. Viral infection may also be an important modulator in IgE-mediated histamine release. Ida et al (26) obtained leukocytes from patients with ragweed allergy, incubated them with respiratory viruses, and then challenged them with a ragweed antigen E. Leukocytes incubated with virus demonstrated a significant enhancement of histamine release, which suggests that exaggerated bronchial reactivity might occur if the patient was exposed to specific antigen during an intercurrent viral infection.

CONCLUSION

The effect of respiratory viral infection on lung function is clearly complex. Key mechanisms seem to be peripheral airway inflammation and exaggerated bronchial reactivity, presumably secondary to epithelial damage and sensitization of airway receptors. From a clinical standpoint, respiratory viral infection may greatly enhance the adverse effects of a variety of physical stimuli including cold air and common pollutants. Further work in this area may help define important host defense mechanisms.

Literature Cited

1. Picken, J. J., Niewoehner, D. E., Chester, E. M. 1972. Prolonged effects of viral infections of the upper respiratory tract upon small airways. *Am. J. Med.* 52:738–46
2. Horner, G. J., Gray, F. D. Jr. 1973. Effect of uncomplicated presumptive influenza on the diffusing capacity of the lung. *Am. Rev. Respir. Dis.* 108:866–69
3. Fridy, W. W., Ingram, R. M., Hierholzer, J. C., Coleman, M. T. 1974. Airways function during mild viral respiratory illnesses. *Ann. Intern. Med.* 80:150–55
4. Collier, A. M., Pimmel, R. L., Hasselblad, V., Clyde, W. A., Knelson, J. H., Brooks, J. G. 1978. Spirometric changes in normal children with upper respiratory infections. *Am. Rev. Respir. Dis.* 117:47–53
5. Macklem, P. T., Mead, J. 1967. Resistance of central and peripheral airways measured by a retrograde catheter. *J. Appl. Physiol.* 22:395–401
6. Hall, W. J., Douglas, R. G. Jr., Hyde, R. W., Roth, F. K., Cross, A. S., Speers, D. M. 1976. Pulmonary mechanics after uncomplicated influenza A infection. *Am. Rev. Respir. Dis.* 113:141–47
7. Little, J. W., Hall, W. J., Douglas, R. G. Jr., Hyde, R. W., Speers, D. M. 1976. Amantadine effect on peripheral airways abnormalities in influenza. *Ann. Intern. Med.* 85:177–82
8. Little, J. W., Hall, W. J., Douglas, R. G. Jr., Mudholkar, G. S., Speers, D. M., Patel, K. 1978. Airway hyperreactivity and peripheral airway dysfunction in influenza A infection. *Am. Rev. Respir. Dis.* 118:295–303
9. Douglas, R. G. Jr. 1979. Respiratory diseases. In *Antiviral Agents and Viral Diseases of Man*, ed. G. J. Galasso, 8:385–459. New York: Raven. 950 pp.
10. Rosenzweig, D. Y., Dwyer, D. J., Ferstenfeld, J. G., Rytel, M. W. 1975. Changes in small airway functions after live attenuated influenza vaccination. *Am. Rev. Respir. Dis.* 111:399–403
11. Zeck, R., Solliday, N., Kehoe, T., Berlin, B. 1976. Respiratory effects of live influenza virus vaccine: Healthy older subjects and patients with chronic respiratory disease. *Am. Rev. Respir. Dis.* 114:1061–67
12. Prevost, J. M., Vereerstraeten-Schmerber, J., Lamy, F., DeKoster, J. D. 1975. Live attenuated influenza virus vaccines in patients with chronic bronchopulmonary diseases. *Scand. J. Respir. Dis.* 56:58–69
13. Little, J. W., Douglas, R. G. Jr., Hall, W. J., Roth, F. K. 1979. Attenuated influenza produced by experimental intranasal inoculation. *J. Med. Virol.* 3:177–88
14. Fell, P. J., Watson, N. P., O'Donnell, H. M., Simmons, R. L., Masell, S. K. 1977. Longer term effects of live influenza vaccine in patients with chronic pulmonary disease. *Lancet* 1:1282–84

15. Douglas, R. G. Jr., Bentley, D. W., Betts, R. F., Zaky, D. A., Roth, F. K., Murphy, B. R. 1976. Evaluation of a temperature-sensitive influenza virus in elderly and chronically ill subjects. *Am. Rev. Respir. Dis.* 113:293–300
16. Walsh, J. J., Dietlein, L. F., Low, F. N., Burch, G. E., Mogabgab, W. J. 1961. Tracheobronchial response in human influenza. *Arch. Intern. Med.* 108: 376–88
17. Empey, D. W., Laitinen, L. W., Jacobs, L., Gold, W. M., Nadel, J. A. 1976. Mechanisms of bronchial hyperreactivity in normal subjects after upper respiratory tract infection. *Am. Rev. Respir. Dis.* 113:131–39
18. Nadel, J. A. 1963. Structure-function relationships in the airway: Bronchoconstriction mediated via vagus nerves or bronchial arteries. *Med. Thorac.* 22:231–42
19. Hall, W. J., Hall, C. B., Speers, D. M. 1978. Respiratory syncytial virus infection in adults. *Ann. Intern. Med.* 88:203–5
20. Aquilina, A. T., Hall, W. J., Douglas, R. G. Jr., Utell, M. J. 1979. Evaluation of pulmonary function abnormalities in viral upper respiratory infections: Effect of exercise and cold air. *Am. Rev. Respir. Dis.* 119:91 (Abstr.)
21. Levy, D., Gent, M., Newhouse, M. T. 1977. Relationships between acute respiratory illness and air pollution levels in an industrial city. *Am. Rev. Respir. Dis.* 116:167–73
22. Utell, M. J., Swinburne, A. J., Hyde, R. W., Speers, D. M., Gibb, F. R., Morrow, P. L. 1979. Airway reactivity to nitrates in normal and asthmatic subjects. *J. Appl. Physiol.* 46:189–95
23. Utell, M. J., Aquilina, A. T., Hall, W. J. 1978. Airway reactivity to nitrates in subjects with naturally acquired influenza A infection. *Chest* 74:355 (Abst.)
24. Szentivanyi, A. 1968. The beta-adrenergic theory of the atopic abnormality in asthma. *J. Allergy* 42:203–32
25. Busse, W. W. 1977. Decreased granulocyte response to isoproterenol in asthma during upper respiratory infections. *Am. Rev. Respir. Dis.* 115:783–91
26. Ida, S., Hooks, J. J., Siraganian, R. P., Notkins, A. L. 1977. Enhancement of IgE-mediated histamine release from human basophils by viruses: Role of interferon. *J. Exp. Med.* 145:892–906

Ann. Rev. Med. 1980. 31:239–45
Copyright © 1980 by Annual Reviews, Inc. All rights reserved

POSTURAL HYPOTENSION

❖7357

Michael G. Ziegler, M.D.

Departments of Medicine and Pharmacology, University of Texas
Medical Branch, Galveston, Texas 77550

Many diseases cause postural hypotension and evaluating them all in every patient whose blood pressure falls upon standing is impossible. However, an examination of the specific mechanisms that maintain blood pressure can quickly point toward the cause of postural hypotension and limit the differential diagnosis to a few diseases. This chapter discusses mechanisms that maintain blood pressure, new tests for their evaluation, and major diseases in which these mechanisms are faulty.

When a person stands, the force of gravity in the vertical axis (+1 Gz) results in pooling of 300–800 ml of blood in the legs (1). After standing for 10 minutes there is a 10% hemoconcentration (2) as water moves out of the vessels into interstitial spaces and causes swelling of the feet and legs. This "self-phlebotomy" reduces central blood volume and filling pressure of the heart. This causes a marked decrease in cardiac output in people lacking compensatory mechanisms, and in those people blood pressure falls to less than half the pressure while recumbent. Normal people, however, have powerful compensatory mechanisms to maintain blood pressure and these remain effective even after the loss of 1000 ml blood or when subjected to a force of +7 Gz for up to one minute.

After standing, normal people increase their heart rate in less than two seconds. This is a neurogenic response initiated by baroceptors located at the arch of the aorta and carotid bifurcation. This baroceptor input to the brainstem is the most important stimulus initiating postural hemodynamic responses. In some people cortical input appears important; they are much more susceptible to hypotension and fainting when tilted upright on a tilt table than when they perform the familiar act of standing.

When a person stands, the sympathetic nervous system increases the activity of nerves to blood vessels and the heart to compensate for a decrease in venous return. Heart rate increases but there is a small decrease in total

0066-4219/80/0401-0239$01.00

cardiac output and a greater decrease in stroke volume. Peripheral resistance increases slightly, which thus maintains mean blood pressure while pulse pressure decreases.

This response to standing depends on several mechanisms; failure of any of them leads to postural hypotension: (*a*) neurogenic feedback loop that depends on baroceptor input to brainstem centers. This inhibits vagal tone and stimulates spinal nerve tracts synapsing with sympathetic ganglia, which in turn stimulate release of norepinephrine (NE) from sympathetic nerves; (*b*) venous and arteriolar vasoconstriction in response to NE; (*c*) cardiac response with positive inotropic and chronotropic effects of NE and a decrease in vagal negative chronotropic effects; (*d*) an adequate blood volume to supply the heart and blood vessels.

The primary neurogenic component of the adaptation to standing is a diffuse increase in sympathetic nervous activity mediated by enhanced release of NE. A tiny fraction of the released NE diffuses into the blood stream and, after standing, blood levels of NE increases by 50% within two minutes, double by five minutes, and triple after mild exercise (Figure 1). Individuals with significant postural hypotension almost always have an abnormal pattern of NE release that can indicate the type of disorder causing hypotension.

Patients with idiopathic orthostatic hypotension, who have isolated autonomic dysfunction, have a marked depletion in NE content of blood vessels normally innervated by sympathetic nerves (3). When these patients are given tyramine, a drug that releases endogenous stores of NE, they fail to respond. While recumbent these patients have plasma NE levels that are only one third of normal, reflecting their depleted tissue stores. When they stand, their blood pressure drops precipitously, they fail to increase their heart rate, and their low level of plasma NE remains unchanged (Figure 1). Even with exercise they cannot increase their circulating NE levels since there is apparently no NE to be released.

Several other diseases can cause destruction of peripheral sympathetic nerves. Thiamine deficiency, alcohol, and other chemicals may cause a neuropathy that includes autonomic nerves. Ganglionic-blocking drugs, reserpine, and guanethidine have a sympatholytic effect. Diabetes is probably the most common cause of autonomic neuropathy; the process develops slowly so that as the diabetes worsens the pattern of NE secretion decreases from a normal response to that characteristic of patients with idiopathic orthostatic hypotension (Figure 1). Diabetics also frequently develop orthostatic hypotension from volume depletion caused by renal salt and albumin wasting and protein leakage through diseased capillaries. When this happens the sympathetic nervous response is greater than normal and is characteristic of the response to volume depletion (Figure 1). An increased NE

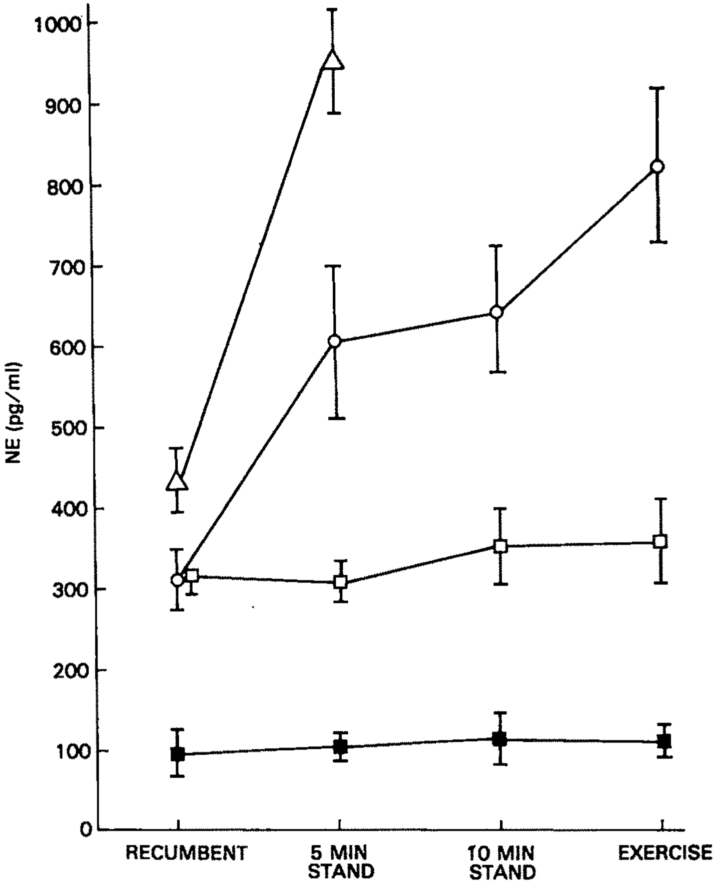


Figure 1 Plasma NE levels while recumbent, after standing for 5 or 10 minutes, and after 5 minutes of isometric exercise while standing. Solid squares represent idiopathic orthostatic hypotension; open squares, Shy-Drager Syndrome; open circles, normal subjects; and open triangles, normal subjects after furosemide diuresis. Values show ± standard error of the mean.

release in response to volume depletion sharply contrasts with the decreased NE levels seen in diabetic autonomic neuropathy and indicates that the patient may benefit from volume expansion (4).

Patients who have intact peripheral sympathetic nerves but who have a defect of the brain stem or spinal cord that prevents activation of these nerves can have severe postural hypotension. Patients with the Shy-Drager Syndrome (5) have abnormalities of their brain stem and postural hypotension. They have a normal response to tyramine infusion (6) and normal resting levels of plasma NE (7). When they stand, however, they fail to activate their sympathetic nervous system, and NE levels remain unchanged (Figure 1) while blood pressure drops. Cerebrovascular insufficiency can present with postural hypotension. Patients who have infarcted vasoregulatory brain areas fail to increase plasma NE on standing. They often have abnormal arteriograms. Other diseases affecting the brain stem and basal ganglia such as Huntington's Chorea (8) and Parkinsonism (9) are often associated with mild postural hypotension and a small decrease in plasma NE from normal levels.

In contrast to patients with defective sympathetic nervous systems, normal subjects double their plasma norepinephrine levels after standing for five or ten minutes and triple their levels after moderate exercise. Normal subjects who have been given sufficient furosemide to cause a weight loss of 1.5 liter of fluid develop some postural hypotension and, while recumbent, have plasma NE levels that are higher than normal and that increase strikingly above normal levels upon standing (Figure 1). This marked increase in sympathetic nervous activity is not quite adequate to maintain their blood pressure while they stand and these subjects decreased their blood pressure from 110/75 to 102/71 while increasing their heart rate from 73 to 88. It is not uncommon to see patients suspected of having neurogenic orthostatic hypotension who are volume depleted. Diabetics with a mild autonomic deficit may also have mild volume depletion and this combination can lead to disabling postural hypotension.

Drugs that prevent the vasoconstriction normally caused by NE have an effect similar to that of volume depletion. When the drug is an α-blocking agent such as phentolamine or a vasodilator such as a nitrite, the postural hypotension provides little diagnostic challenge. However, other α-blocking drugs such as chlorpromazine or vasodilators such as alcohol can have similar effects. When these drugs do not effect a β-adrenergic blockade, the tachycardia in response to standing may serve as an important clue to their presence.

Postural dizziness may be the presenting complaint of patients with a central or peripheral nerve defect or with hypovolemia. These groups of patients can be distinguished on the basis of their patterns of NE secretion and by physical examination. Central nervous system disease that incapacitates the sympathetic nervous system almost always involves adjacent structures. Interruption of spinal pathways to sympathetic nerves is always associated with other signs of a cord lesion, usually paralysis and spasticity. Involvement of the vasomotor centers is usually accompanied by movement disorders, Parkinsonism, bulbar defects, weakness, or spasticity. These should all be tested for with the patient recumbent since hypotension can lead to transient neurologic deficits. Patients with degeneration of peripheral autonomic nerves may develop central defects coincidentally or as the result of episodes of hypotension, but such cases are not common.

Diseases causing neuropathic changes in peripheral sympathetic fibers depress both resting and stimulated levels of plasma NE since they destroy neuronal stores of NE. The disease may affect only autonomic nerves as in idiopathic orthostatic hypotension, may affect a wide range of peripheral nerves as in diabetic and toxic neuropathies, or may affect peripheral and central nerves as in the Guillain-Barré syndrome. All these diseases lower catecholamine levels in tissues innervated by sympathetic nerves and decrease plasma NE at rest and while standing. Patients with these diseases have decreased responsiveness to infusions of tyramine, which release endogenous NE, and increased responsiveness to infusions of NE, due to denervation supersensitivity.

All the diseases of both central and peripheral vasoregulatory nerves that cause postural hypotension affect both sympathetic and parasympathetic nerves. As a result, patients with these diseases present a constellation of symptoms that are very similar in diseases of diverse etiology. They may have a fixed heart rate, incontinence, constipation, inability to sweat, heat intolerance, and fatigability. In contrast, patients with postural hypotension due to volume depletion, venous pooling of blood, or vasodilator drugs have a marked increase in pulse rate on standing and do not have signs of autonomic deficiency. However, it is not uncommon for diabetics with postural hypotension to have both mild autonomic neuropathy and volume depletion. Diagnosis of this combination requires measurement of plasma NE levels, and sometimes of plasma and red blood cell volumes.

Successful therapy of postural hypotension depends on a precise diagnosis. The most common cause is iatrogenic when a combination of volume depleting and sympatholytic drugs are given to treat hypertension. Although guanethidine is the best known cause of iatrogenic postural hypotension, drug combinations that act on both central and peripheral sympathetic nervous control (clonidine, methyldopa, prazosin), on cardiac β receptors (propranolol, metoprolol), and on plasma volume (diuretics) attack all the major means of postural blood pressure control and can lead to marked postural hypotension. Adequate blood pressure control, without disabling postural hypotension, can often only be achieved if one of the sympatholytic agents is replaced with an arterial vasodilator such as hydralazine or minoxidil.

Patients with postural hypotension due to volume depletion alone usually have low blood pressure while recumbent and respond well to volume expansion through replacement of plasma protein, red cells, or fluid with fludrocortisone and supplementary sodium chloride.

Patients with neurogenic hypotension due to central or peripheral causes are difficult to treat adequately. Their blood pressure can be raised with α adrenergic agonists such as phenylephrine, hydroxyamphetamine, or, in patients with a central defect, a monoamine oxidase inhibitor and tyramine to release endogenous NE. These can have pressor effects and yet not provide symptomatic relief since blood flow to the brain may be impeded by these agents, which cause diffuse vasoconstriction. Agents that stimulate the heart, such as ephedrine, may provide more symptomatic relief provided the patient has an adequate return of blood to the heart. Since patients with neurogenic postural hypotension cannot adequately constrict their venous capacitance vessels when they stand, treatment can be aimed at compensating for this defect. Fitted Jobst® stockings that extend waist high, or an anti-G suit, can decrease venous pooling. Administration of fludrocortisone and salt is usually better accepted by patients but can cause hypertension and congestive heart failure during recumbency. Propranolol, to block β vasodilation, has been helpful in some people and indomethacin or flurbiprofen, to block prostaglandin induced vasodilation (10, 11), may help but these drugs are not uniformly successful (12).

Sometimes treatment of a disease may unmask a latent defect in postural responses. For example, some symptoms of the Shy-Drager syndrome and Parkinsonism respond to levodopa or bromocriptine, but these drugs aggrevate postural hypotension (13). Treatment of fluid overload in renal failure may unmask the hypotensive effects of uremic autonomic neuropathy (14). Insulin can have unexpected hypotensive effects in patients with mild diabetic autonomic neuropathy (15). Such drug responses can serve as important clues to underlying disease.

Diseases or drugs that cause postural hypotension usually impair (*a*) peripheral autonomic nerves; (*b*) central control of autonomic nerves, or (*c*) blood volume. Hypotensive patients can usually be categorized into one of these three groups; this limits the differential diagnosis and can serve as a guide to effective therapy.

Literature Cited

1. Sjöstrand, T. 1952. Regulation of the blood distribution in man. *Acta Physiol. Scand.* 26:312
2. Ziegler, M. G., Lake, C. R., Kopin, I. J. 1976. Deficient sympathetic nervous response in familial dysautonomia. *N. Engl. J. Med.* 294:630–33
3. Kontos, H. A., Richardson, D. W., Norvell, J. E. 1975. Norepinephrine depletion in idiopathic orthostatic hypotension. *Ann. Intern. Med.* 82:336–41
4. Cryer, P. E., Silverberg, A. B., Santiago, J. V., Shah, S. D. 1978. Plasma catecholamines in diabetes: The syndromes of hypoadrenergic and hyperadrenergic postural hypotension. *Am. J. Med.* 64:407–16
5. Shy, G. M., Drager, G. A. 1960. A neurological syndrome associated with orthostatic hypotension: a clinical-pathologic study. *Arch. Neurol.* 2: 511–27
6. Kontos, H. A., Richardson, D. W., Norvell, J. E. 1975. Mechanisms of circulatory dysfunction in orthostatic hypotension. *Trans. Am. Clin. Climatol. Assoc.* 87:26–35
7. Ziegler, M. G., Lake, C. R., Kopin, I. J. 1977. The sympathetic-nervous-system defect in primary orthostatic hypotension. *N. Engl. J. Med.* 296:293–97
8. Shoulson, I., Ziegler, M. G., Lake, C. R. 1976. Huntington's Disease (HD): Determination of plasma norepinephrine (NE) and dopamine-beta-hydroxylase (DBH). *Neurosci. Abstr.* 2:800
9. Teychenne, P. F., Lake, C. R., Ziegler, M. G., Plotkin, C., Wood, J. H., Calne, D. B. 1977. Central and peripheral deficiency of norepinephrine in Parkinson's Disease and the effects of L-dopa therapy. *Neurosci. Abstr.* 3:417
10. Perkins, C. M., Lee, M. R. 1978. Flurbiprofen and fludrocortisone in severe autonomic neuropathy. *Lancet* 2:1058
11. Kochar, M. S., Itshovitz, H. D. 1978. Treatment of idiopathic orthostatic hypotension (Shy-Drager Syndrome) with indomethacin. *Lancet* 1:1011
12. Bannister, R., Davies, B., Sever, P. 1978. Indomethacin for Shy-Drager Syndrome. *Lancet* 1:1312
13. Ziegler, M. G., Lake, C. R., Williams, A. C., Teychenne, P. F., Shoulson, I., Steinsland, O. 1978. Bromocriptine inhibits norepinephrine release. *Clin. Pharmacol. Ther.* 25(2):137–42
14. Ziegler, M. G., Lake, C. R., Coleman, M., Kopin, I. J. 1977. Sympathetic nerve defect in chronic renal failure. *Am. Soc. Nephrol. Abstr., 10th Ann. Meet.*, p. 42.
15. Page, M. M., Watkins, P. J. 1975. Provocation of postural hypotension by insulin in diabetic autonomic neuropathy. *Diabetes* 25(2):90–95

Ann. Rev. Med. 1980. 31:247–59
Copyright © 1980 by Annual Reviews Inc. All rights reserved

SNAKE VENOM POISONING IN THE UNITED STATES

❖7358

Findlay E. Russell, M.D., Ph.D.[1]

Departments of Neurology, Physiology, and Biology, University of Southern
California, School of Medicine, Los Angeles County–University of Southern
California Medical Center, Los Angeles, California 90033

INTRODUCTION

Snake venom poisoning is a medical emergency that requires immediate attention and the exercise of considerable judgment. It is a multiple type of poisoning, for snake venoms are complex mixtures containing as many as ten biologically active components. It must be remembered, however, that in attending a patient who has been bitten by a venomous snake it does not necessarily follow that he has been envenomated by the snake. Approximately 20% of all bites by venomous snakes in the United States do not result in poisoning. The snake may not eject venom or, if it does, it is spewed on the skin in a superficial bite. The physician should establish the diagnosis of snake venom poisoning on the basis of the clinical findings, rather than on the presence of fang or teeth marks and the identification of the offending reptile. Although there are an estimated 45,000 bites by all snakes in the United States each year (1), only about 6680 persons are treated for snake venom poisoning (2). However, it can be expected that at least 1000 additional bites by venomous snakes occur each year and that they are either not treated or go unreported. During the past five years the number of deaths from snakebite in the United States has ranged between 9 and 14. Most of these deaths occurred in children, in the elderly, in untreated, mistreated, or undertreated cases, in cases complicated by other serious

---

[1] Studies on snake venom poisoning are supported in part by grants 1 RO1 GM 24141–01 from the NIGMS and 5 RO1 NS11476–05 from the NINCDS, and by the Gould Family Foundation, Inc.

0066-4219/80/0401-0247$01.00

disease states, or in members of religious sects who handle serpents as part of their worship exercises and refuse medical treatment. Almost all reported deaths have been attributed to the rattlesnakes (3).

DISTRIBUTION

There are approximately 120 species of snakes in the United States of which 20 can be considered sufficiently venomous to be dangerous to man. These poisonous snakes are generally divided into two groups: the pit vipers (Crotalidae) and the coral snakes (Elapidae). The crotalids include the rattlesnakes, cottonmouths (water moccasins), copperheads, the pigmy rattlesnakes, and the massasaugas, while the elapids are the eastern coral snake, and the smaller Arizona or Sonoran coral snake. The distribution of these snakes is shown in Table 1.

In addition to these snakes, bites by "exotic snakes" (those not native to the United States) are being reported with increased frequency. The increase in the number of such snakes in collections and the number of reported bites has been remarkable. It is not possible in this review to discuss the problem of envenomation by the exotic snakes. The reader is referred elsewhere for a thorough discussion of their bites (3–5). Should the physician be called upon to treat such a case, he should contact the local zoo or the Oklahoma City Poison Information Center (303–629–1123), which, in cooperation with the Oklahoma City Zoo and the American Society of Zoological Parks, maintains an *Antivenin Index* listing the location of all foreign antivenins, specialists in snake venom poisoning, and related information. Because of federal regulations regarding the use of foreign-prepared antivenins, it is wise to contact a local public health officer before administering these antisera.

SYMPTOMS AND SIGNS

Pit Viper Envenomation

The more common symptoms and signs of rattlesnake bite are shown in Table 2. Bites by the rattlesnakes, cottonmouths, and copperheads cause some immediate swelling and edema, and varying degrees of local pain. Contrary to popular opinion, severe pain is not a constant finding, and pain may even be absent following poisoning by some species. However, severe pain may be associated with increased swelling and edema. When there is an envenomation, some swelling is usually observed within ten minutes following the bite. It is most unusual if its onset is delayed more than 20 minutes. By the time the patient arrives at the doctor's office, which in most cases in the US varies between 15 and 60 minutes, a diagnosis of rattlesnake bite, with or without envenomation, can usually be made on the basis of fang marks, the presence or absence of edema and pain, and, an important consequence of bites by some species, perioral tingling or numbness with a metallic, minty, or rubbery taste in the mouth. Fine muscle fasciculations may also be present following some rattlesnake venom poisonings.

In moderate or severe envenomations, the edema progresses rapidly and may involve the entire extremity within a few hours. There may be lymphangitis, and regional lymph nodes are often enlarged and tender. Weakness, syncope, sweating, and nausea may be present. If the envenomation is more severe, there may be paresthesia over the scalp and at the finger and toe tips, particularly following bites by the eastern diamondback, timber, and Pacific rattlesnakes. Vomiting may occur; the pulse may be rapid and weak; blood pressure is usually below normal; and, in severe poisonings, shock may develop early. Respiratory distress may be present, particularly following bites by the Mojave and eastern diamondback rattlesnakes. Muscle fasciculations, spasms, and weakness are not uncommon; true paralysis may occur in rare cases. The patient may complain of headache, blurred vision, and marked thirst.

Ecchymosis is a constant finding in almost every untreated case of moderate or severe rattlesnake venom poisoning and often appears over the injured part within six hours of the bite. It is most severe following bites by the eastern and western diamondbacks, and the prairie and Pacific rattlesnakes, and least severe following copperhead and Mojave rattlesnake bites. The skin may appear tense and discolored, and, in untreated or undertreated bites, vesicles often appear within the first eight hours, particularly in the area of the bite. These vesicles may become filled with blood. In general, these changes tend to be relatively superficial in untreated and undertreated cases. There may be bleeding phenomena, as reflected in hemorrhage from the gums, hematemesis, melena, and hematuria. Bleeding and clotting times are prolonged and platelet counts may fall sharply in moderate or severe envenomations. In most cases, there is hemoconcentration early in the course of the poisoning, and in moderate or severe cases a rapid fall in hematocrit and hemolysis may subsequently occur.

Coral Snake Envenomation

The bite is usually associated with some pain, although it may be minor and transitory in nature. Swelling is either absent or very minor. Paresthesia is sometimes noted around the bitten area, and some weakness of the part may become evident within several hours of the poisoning. The patient may complain of drowsiness, apprehension, and weakness. Muscular incoordination may develop, and muscle fasciculations and tremor of the tongue may be seen. Increased salivation and difficulties in swallowing and phonation, as well as visual disturbances, respiratory distress and failure, a bulbar type of paralysis, convulsions, and shock may develop.

Table 1 States in which some medically important snakes occur[a]

	Wash., Ore., Ida.	Calif., Nev.	Ariz., N. Mex.	Tex.	Mont., Mich., Wisc., Minn., S. Dak., N. Dak., Nebr., Iowa, Utah, Colo.	Kans., Okla., Ark., Mo.	Tenn., Ky., Ill., Ind., Ohio	N.C., S.C., Ga., Ala., Miss., La.	Fla.	Penn., N.J., Md., Del., Va., W. Va., N.Y., N. Engl.
ELAPIDAE										
Eastern coral snake (*Micrurus fulvius*)				X		Ark.		X	X	
Western coral snake (*Micruroides euryxanthus*)			X							
CROTALIDAE										
Cottonmouths and Copperheads (*Agkistrodon*)										
Cottonmouths (*A. piscivorus*)				X	Nebr., Iowa	X	Tenn., Ky., Ill.	X	X	Va.
Copperheads (*A. contortrix*)				X	Nebr., Iowa	X	X	X		X
Rattlesnakes (*Crotalus*)										
Eastern diamondback (*C. adamanteus*)								X	X	
Western diamondback (*C. atrox*)		X	X	X		Okla., Ark.				
Sidewinder (*C. cerastes*)		X	Ariz.		Utah					
Timber (*C. horridus*)				X	Minn., Wisc., Iowa, Nebr.	X	X	X	X	X
Rock (*C. lepidus*)			X	X						
Speckled (*C. mitchelli*)		X	Ariz.							

Black-tailed (*C. molossus*)			X	X						
Twin-spotted (*C. pricei*)			Ariz.							
Red diamond (*C. ruber*)		X								
Mojave (*C. scutulatus*)		X	X	X						
Tiger (*C. tigris*)			Ariz.							
Western (*C. viridis*) Prairie (*C.v. viridis*)	Ore., Ida.		X	X	Not Mich., Minn., Wisc.	Kans., Okla.				
Grand Canyon (*C.v. abyssus*)			Ariz.							
Southern Pacific (*C.v. helleri*)		Calif.								
Great Basin (*C.v. lutosus*)	Ore., Ida.	X	Ariz.		Utah					
Northern Pacific (*C.v. oreganus*)	X	X								
Ridge-nosed (*C. willardi*)			Ariz.							
Massasauga and Pigmy (*Sistrurus*) Eastern massasauga (*S. catenatus*)			X	X	Mich., Wisc., Minn., Nebr., Iowa, Colo.	X	Ill., Ind., Ohio			N.Y., Pa.
Pigmy rattlesnake (*S. miliarius*)				X		Okla., Ark., Mo.	Tenn.	X	X	

[a] Certain groups of adjoining states are here treated as units. The symbol X indicates distribution of the species is widespread within the unit. Restriction of a species to a part of a unit is indicated appropriately.

Table 2 Frequency of symptoms or signs of rattlesnake venom poisoning[a]

Fang marks	100	Tingling or numbness of affected part	57
Swelling and edema	74	Necrosis	27
Pain	65	Respiratory rate changes	40
Ecchymosis	51	Decreased hemoglobin	37
Vesiculations	40	Abnormal electrocardiogram	26
Changes in pulse rate	60	Cyanosis	16
Weakness	70	Hematemesis, hematuria, or melena	15
Sweating and/or chill	43	Glycosuria	20
Numbness or tingling of tongue and mouth, or scalp or feet	63	Proteinuria	16
Faintness or dizziness	57	Unconsciousness	12
Nausea, vomiting, or both	42	Thirst	34
Blood pressure changes	54	Increased salivation	20
Change in body temperature	42	Swollen eyelids	7
Swelling regional lymph nodes	40	Retinal hemorrhage	5
Fasciculations	41	Blurring of vision	12
Increased blood clotting time	37	Convulsions	1
Sphering of red blood cells	18	Muscle contractions	6
		Increased blood platelets	4
		Decreased blood platelets	42

[a] Total number of cases = 100 for all entries except "Increased blood platelets," where total number of cases = 25.

TREATMENT

First Aid

Far too much importance has been placed on the first aid treatment of pit viper and coral snake envenomation at the expense of definitive medical care. Other than (*a*) putting the victim at rest and keeping him warm, (*b*) giving him reassurance, (*c*) immobilizing the affected part in a functional position at heart level, (*d*) watching for any untoward reaction, and (*e*) transporting him to the nearest medical facility as quickly as possible, further first aid may be of questionable value. There is no danger in placing a constriction band, not a tourniquet, immediately proximal to the wound in any case of snakebite. It may help to retard lymphatic flow. If the patient is more than 20–30 minutes from medical care, and is seen within five minutes of the bite of a rattlesnake or cottonmouth, place a constriction band immediately proximal to the wound and incise through each fang mark. Incisions should be no more than 1 cm in length and no deeper than 3 mm. Apply suction directly over both incisions. This procedure is of no value if delayed more than 15 minutes following the bite. Suction should be carried out for 30 minutes to one hour. An attempt should be made to identify the snake but no time should be wasted in this act. After immobilizing the affected part, the victim should then proceed to the nearest medical facility or, if possible, be taken there at rest. Under no conditions should the injured part be placed in ice, the bite area excised, or surgical measures performed, unless a physician is present and deems such a procedure necessary. In cases of coral snake bite the injured part should be loosely immobilized and the patient rushed to the nearest medical facility.

Laboratory Tests

On admission to the hospital, and in the presence of obvious findings of envenomation, the necessary laboratory tests should be done. The choice of tests will depend upon the kind of snake involved. In cases of rattlesnake envenomation, typing and cross-matching; bleeding, clotting and clot retraction times; complete blood and platelet counts; and urinalysis should be done immediately. We frequently obtain red blood cell indices, sedimentation rate, prothrombin time, arterial blood gas measurements, and determinations of sodium, potassium, and chloride. In severe poisonings, it is advisable to get an electrocardiogram. We obtain them on all patients moderately or severely envenomated. Serum protein levels, fibrinogen titer, partial thromboplastin time, and renal function tests are often useful. In severe envenomations by rattlesnakes, the hematocrit value, blood cell count, hemoglobin concentration, and platelet count should be obtained several times a day for the first few days, and all urine samples should be examined, particularly for red blood cells. In cases of coral snake bite, emphasis on the blood screens is not necessary (6).

Medical Management

ANTIVENIN On admission to the medical facility, if envenomation has occurred and the use of antivenin seems warranted, a skin test for sensitivity to horse serum should be performed, following the instructions in the accompanying antivenin brochure. For crotalid bites, the amount of Antivenin [Crotalidae] Polyvalent to be given will depend on many factors, most important being the severity of the symptoms and signs and their progression. I do not advise the use of antivenin for very minor envenomations. When at 30 to 60 minutes the swelling is confined to the area of the bite, there are no paresthesias, fasciculations, or ecchymosis, and no symptoms other than very minor pain, and if laboratory findings are normal, it is not necessary to give antivenin. However, such patients should be observed for four hours. The wound area should be cleansed and the affected part immobilized at heart level. We have done this in more than 40 patients without subsequent troubling reactions or complications, except for minor wound infections in three cases. Following discharge, the patient should remain in bed at home and limit his physical activities for at least another

24 hours. He should be advised to keep his hand, or foot, elevated for several days. It is also advisable to give the appropriate antitetanus agent.

Antivenin is indicated in those patients who, within 30 to 60 minutes following the bite, show progressive swelling involving the injured area, who complain of paresthesia about the mouth or over the scalp or at the finger or toe tips, and who have any systemic symptoms or signs of poisoning. In such cases, the girth or circumference of the finger, hand, toe, or foot and two or more sites proximal, such as at the wrist and upper forearm, or ankle and thigh, should be measured. These measurements can be repeated every 15 minutes. They serve as an index of the progression of the poisoning, as well as a guide for antivenin administration.

In minimal envenomations, three to five vials or units of antivenin, intravenously, will usually suffice. The antivenin should be mixed with the dilutant in the vial in which it is supplied, and then put in 500 ml of intravenous fluid. In minimal envenomations, administer the first three to five vials within one hour and if swelling does not progress and paresthesias disappear (which does not always happen), start a second 500 ml of fluid with several additional vials of antivenin. This can be administered over the next several hours. If the swelling progresses at any time during this period, add several more vials of antivenin to the drip. In such cases, of course, the envenomation was probably a moderate rather than a minimal one. It is wise to piggy-back 1000 ml of 5% glucose in water and epinephrine on the same intravenous set-up, in the event of a serum reaction, although such reactions are very rare. Benadryl® should be kept at hand, for in an occasional patient the too rapid infusion of antivenin may cause a reaction. This usually responds to slowing the infusion rate and/or injecting Benadryl.

In moderate envenomations, where the edema has invaded or encompassed the hand or foot within 30 minutes of the bite and is progressing; and where there is mild cyanosis or ecchymosis of the area; and the patient complains of perioral paresthesia and tingling of the scalp, or finger or toe tips, feels nauseated or weak; and is hypotensive, or presents other systemic signs, six to ten vials of antivenin should be given in 500 ml of the i.v. drip during the first hour. If the manifestations worsen under this regimen, then the patient may have suffered a severe envenomation and require another 5–20 vials of antivenin, or even more. In such cases, a high level of antivenin should be maintained during the first four hours following the bite. The amount of dilutant will need to be consistent with the electrolyte balance, renal function, and circulating blood volume.

In severe poisonings, a minimum of ten vials of antivenin is necessary in the first i.v. drip. If the symptoms and signs are lessened, or their progression is considerably slowed by this initial dose, prepare another five units and give this over a two- to four-hour period. If the manifestations are not affected by the first ten-vial infusion, or only slightly so, give 5–10 additional units over the second hour. It may be necessary to give 20 or 30 vials of antivenin. Since some of the patients may be hypotensive or in shock, other measures beside the injection of antivenin will be needed. Shock may not respond to antivenin alone since the hypotension is caused by the loss of circulating fluid with decrease in blood volume, a condition that antivenin cannot restore. In such cases, shock measures may need to be instituted.

The route of Antivenin [Crotalidae] Polyvalent administration has been reviewed by the manufacturer. The effectiveness of intravenous antivenin was first demonstrated by our group and by McCullough & Gennaro at the 1960 meeting of the Ad Hoc Committee on Snakebite Therapy, of the National Academy of Sciences. Since then the efficacy of the intravenous route has been demonstrated in both man and animals (7–9). It is the route of choice, since adequate blood levels are quickly obtained. In the severely envenomated patient, who is also likely to be in shock, intramuscular injection of the antivenin will not give a sufficiently high titer of antibodies to be effective during the first six or eight hours.

Antivenin should never be injected into a finger or toe. Some physicians inject one-third of one vial into the subcutaneous tissues in the area of the bite, when it is not on a finger or toe. It is not known how long after envenomation antivenin can be given and still be effective. Antivenin [Crotalidae] Polyvalent is most effective if given within four hours of the bite. It is of less value if delayed for eight hours, and of questionable value after 24 hours, although it appears to have been given after that time with a good response. Certainly, there is some clinical and experimental work to indicate that antivenin can be effective when administered beyond the 24-hour post-bite period. The antivenin neutralizes both the lethal and local tissue effects of the venom (3).

In coral snake bites, if any symptoms or signs of envenomation develop during the first several hours following the bite, three vials of Antivenin (*Micrurus fulvius*) should be given intravenously. However, if a patient has definitely been bitten, and particularly if there is some initial pain, five vials of antivenin should be given as soon as possible after the bite. In times past there has been a tendency to wait until systemic manifestations developed before administering antivenin. Experience has shown that once symptoms and signs appear, it may not always be possible to prevent further action of the venom or reverse those changes that have already occurred. It seems wisest to begin antivenin treatment as soon as possible after the bite. No doubt this may be overtreating some or even most patients, but the alternative is far less desirable.

INTRAVENOUS FLUIDS A decrease in circulating blood volume and perfusion failure is a common finding in all severe and most moderate rattlesnake venom poisonings. There may also be lysis of red blood cells and loss of platelets, necessitating transfusions and parenteral fluids. In patients with hypovolemia, plasma or albumin may be used to restore circulating blood volume (3, 10). If there is evidence of a decrease in red blood cell mass, either from lysis of the cells or bleeding, packed blood cells or whole blood should be given. Where these complications are accompanied by defects in hemostasis, including coagulopathies, then replacement with specific clotting factors, fresh whole blood, plasma, or platelet transfusions may be indicated.

ANTIBIOTICS Most physicians give antibiotics routinely following snake bite. Certainly, where there is severe tissue involvement a broad-spectrum antimicrobial agent should be given. In very minor or minimal envenomations antibiotics may not be necessary, although osteomyelitis has developed following at least one case of snake venom poisoning in which an antibiotic was not used (3).

ANTITETANUS AGENTS The appropriate antitetanus agent should always be given.

ANALGESICS In most cases the severe pain associated with snakebite is related to the degree of swelling and edema. The initial pain produced by the venom is seldom marked, the one exception being bites by the eastern diamondback. In most cases, acetylsalicylate, acetaminophen, or codeine are sufficient to control pain. Meperidine hydrochloride may be necessary but this drug is likely to provoke more than the usual incidence of vomiting in patients with venom poisoning, which in the presence of a tendency toward gastrointestinal bleeding is undesirable. The use of morphine should be avoided.

SEDATION Mild sedation is indicated in all moderate or severe poisonings where respiratory depression is not a problem. Five milligrams of Valium,® morning and evening, is usually sufficient to keep the patient hypoactive. This can be supplemented for sleep. Do not give sedatives to patients envenomated by the Mojave rattlesnake or the coral snakes.

ANTIHISTAMINES These are of no value in acute poisoning but are used by many physicians with the thought of suppressing or forestalling sensitivity reactions to the horse serum. Their use, however, is not contraindicated.

DIET After the first 24 hours and in the absences of nausea, vomiting, mouth or throat paresthesia, bleeding problems, or coagulation defects, the patient can be placed on a soft diet. During the first 24 hours the patient's diet should be restricted to liquids, unless the envenomation has been a relatively minor one.

RESPIRATORY FAILURE At the first signs of respiratory depression, which is infrequently seen, oxygen should be initiated and preparations made for assisted or controlled respirations. Tracheal intubation may be necessary but tracheostomy should be avoided. In any severely envenomated patient who is hypotensive, respiratory assistance will be of value, while oxygen may be unnecessary.

RENAL FAILURE The routine measures for the treatment of acute renal failure should be instituted if this finding develops. Renal dialysis is of value in the presence of uremia. It is of questionable value in removing crotalid venom components, since some of the more biologically active fractions have a relatively high molecular weight. Peritoneal dialysis has not proven useful in my experience.

EXCHANGE TRANSFUSIONS These are of value in cases when there are serious coagulation and bleeding deficits.

STEROIDS Most experimental studies show that steroids have no place in the acute phase of snake venom poisoning (11, 12). They may even be contraindicated, in light of the number of deaths occurring in cases where they have been employed. In the only controlled clinical study on their use, Reid et al concluded that, "Prednisone seems to benefit neither systemic nor local poisoning" (13). Visser & Chapman (4), in their excellent text, *Snakes and Snakebite,* state: "There is so much evidence of grave danger from the use of steroids, that they must be avoided in snakebite treatment . . . their only place is in the control of antivenom hypersensitivity and serum reactions."

HEPARIN The use of heparin has been suggested for almost every kind of snake venom poisoning. Both experimental and clinical evidence would seem to indicate its uselessness and even danger.

FASCIOTOMY This is practiced by some physicians in the US but has few advocates abroad (3). Our group has not had to resort to this measure in the more than 650 cases of snake venom poisoning we have treated. The use of fasciotomies usually reflects an insufficient dosage of antivenin, or no antivenin, during the first 12 hours following the bite. Its rationale has in no way been made clear. Experiments in dogs and observations in humans have shown that the incidence and degree of necrosis in edematous limbs, whether the edema be subcutaneous or in a muscle compartment, is not altered by fasciotomy (14). The use of this measure in all cases of crotalid envenomation, as has been advocated (15), seems ill advised (16). In my opinion, fasciotomies are an unnecessary valueless, and expensive form of immediate therapy. In a hemorrhagic state they may be dangerous and they sometimes lead to unnecessary disfigurement. There may be some indication for their use in a rare case but their routine use is unacceptable.

SUPPORTIVE AND FOLLOW-UP CARE Any wound area should be cleansed and covered with a sterile dressing, but it should not be tightly bandaged. The extremity should be immobilized in a splint and in a functional position. The injured part should be kept at heart level during the first 24 hours. If indicated, daily cleansing of the wound with several 15-minute soaks in 1:20 Burow's solution, and painting of the wounds with an aqueous dye containing brilliant green 1:400, gentian violet 1:400, and acriflavin 1:1000 three to four times weekly is beneficial. Several exposures of affected areas to a continuous oxygen flow, with the arm or leg in a plastic bag, are of value. Surgical debridement, if necessary, should be done between the third and sixth day. The patient should be examined for joint motion, muscle strength, sensation, and girth measurements. Passive and then active physical therapy exercises should be initiated on the third or fourth post-bite day (17). Currently, most of our patients are discharged from the medical center on the second or third day following the bite, and the development of lesions requiring extensive care is rare. The time in hospital has been greatly reduced by the use of larger doses of antivenin than were employed a decade ago.

There are a number of special considerations that may influence snake venom poisoning and its management not reviewed here. These include the problem and treatment of shock, skin testing and procedures to use in the event of a positive skin test, reactions to antivenins, treatment of patients sensitive to snake venom, immunodiagnosis, and the pathophysiology of snake venom poisoning. The interested reader will find reviews of these subjects elsewhere (3, 4).

Literature Cited

1. Russell, F. E. 1968. Snakebite. In *Current Therapy,* ed. H. F. Conn, pp. 830–32. Philadelphia: Saunders
2. Parrish, H. M. 1966. Mortality from snakebites, United States, 1950–54. *Public Health Rep.* 72:1027–30
3. Russell, F. E. 1979. *Snake Venom Poisoning.* Philadelphia: Lippincott. 302 pp.
4. Visser, J., Chapman, D. S. 1978. *Snakes and Snakebite.* Cape Town: Purnell. 152 pp.
5. Reid, H. A. 1972. Snakebite. Part 2. Treatment. *Trop. Doctor* 2:159–63
6. Russell, F. E. 1978. Poisonous snakes, Part V. Snake venom poisoning in the United States. The clinical problem. *Clin. Med.* 85:21–30
7. Parrish, H. M. 1963. Intravenous antivenin in clinical snake venom poisoning. *Mo. Med.* 60:240–44
8. Minton, S. A. Jr. 1974. *Venom Diseases.* Springfield, Ill: Thomas. 235 pp.
9. Russell, F. E., Ružić, N., Gonzalez, H. C. 1973. Effectiveness of Antivenin (Crotalidae) Polyvalent following injection of *Crotalus* venom. *Toxicon* 11: 461–64
10. Schaeffer, R. C. Jr., Carlson, R. W., Puri, V. K., Callahan, G., Russell, F. E., Weil, M. H. 1978. The effects of colloidal and crystalloidal fluids on rattlesnake venom shock in the rat. *J. Pharmacol. Exp. Ther.* 206:687–95
11. Schottler, W. H. A. 1954. Antihistamine, ACTH, cortisone, hydrocortisone and anesthetics in snake bite. *Am. J. Trop. Med. Hyg.* 3:1083–91
12. Allam, M. W., Weiner, D., Lukens, F. D. W. 1956. Comparison of cortisone and antivenin in the treatment of crotaline envenomation. In *Venoms,* ed. E. Buckley, N. Porges, pp. 393–97. Washington, DC: Am. Assoc. Adv. Sci.
13. Reid, H. A., Thean, P. C. Martin, W. J. 1963. Specific antivenene and prednisone in viper-bite poisoning: controlled trial. *Br. Med. J.* 2:1378–80
14. Garfin, S. R., Castilonia, R. R., Mubarak, S. J., Hargens, A. R., Akeson, W. H., Russell, F. E. 1980. A clinical and laboratory evaluation of the role of surgical decompression in the treatment of rattlesnake bites. *Am. Acad. Orthopedic Surg.* Feb.
15. Glass, T. G. Jr. 1976. Early debridement in pit viper bites. *J. Am. Med. Assoc.* 235:2513–16
16. Reid, H. A. 1976. Venoms and antivenoms. (Review on Glass, T. G. Jr. 1976. Early debridement in pit viper bites. *J. Am. Med. Assoc.* 235:2513–16.) *Abstr. Hyg.* 51:1139
17. Jenkins, M., Russell, F. E. 1974. Physical therapy for snake venom poisoning. *Phys. Ther.* 54:1298–1300

Ann. Rev. Med. 1980. 31:261–72
Copyright © 1980 by Annual Reviews Inc. All rights reserved

EFFICACY AND SAFETY OF ORAL HYPOGLYCEMIC AGENTS

❖7359

Holbrooke S. Seltzer, M.D.

Veterans Administration Hospital and Department of Internal Medicine,
The University of Texas Southwestern Medical School, Dallas, Texas 75235

INTRODUCTION

Since 1956 oral hypoglycemic agents have been widely used in the United States for the management of maturity-onset diabetes. All but one of them are sulfonylurea compounds, four of which are in clinical use and two others have undergone extensive clinical trials. All sulfonylureas have the same mode of hypoglycemic action, and maximal doses of each have approximately the same potency for lowering blood sugar. One other hypoglycemic agent, the non-sulfonylurea phenformin, was also used extensively from 1956 until 1977, when it was banned by a federal ruling as an "imminent hazard" to health. The present discussion therefore deals almost entirely with sulfonylureas.

The two main questions regarding the use of sulfonylureas are (*a*) do they work and (*b*) are they safe? The first question concerns their efficacy in controlling hyperglycemia during both short-term and long-term use. The second question concerns both the frequency of sulfonylurea-induced hypoglycemia and their alleged cardiotoxicity. Discussion of the latter entails a review of the so-called UGDP controversy.

CLINICAL PHARMACOLOGY OF SULFONYLUREAS

Evaluation of the efficacy of sulfonylureas should be based upon their appropriate use in properly selected patients.

Agents Available

Table 1 shows the four "first generation" sulfonylureas (tolbutamide, chlorpropamide, acetohexamide, and tolazamide), which have been in clinical use from 1956 to 1966 (1); it also shows two "second generation" agents

0066-4219/80/0401-0261$01.00

Table 1 Sulfonylureas used in the United States

Generic name	Trade name	Daily dosage (mg)
First generation		
Tolbutamide	Orinase®	500–3000
Chlorpropamide	Diabinese®	100–500
Acetohexamide	Dymelor®	250–1500
Tolazamide	Tolinase®	150–750
Second generation		
Glyburide	Glibenclamide® Micronase®	2.5–20
Glipizide	Glucotrol®	2.5–40

(glyburide and glipizide), which have been extensively evaluated in the US but have not been approved for general use. The dosage range for each drug is also given, from the usual starting dose to the maximal dose for each agent. Tolbutamide is the shortest acting agent (about 6 hours) and is given in divided doses. Chlorpropamide is the longest acting drug (about 36 hours) and can be given as a single daily dosage. Acetohexamide and tolazamide are intermediate acting (about 12–18 hours) and are usually given in divided doses. The maximal dosages of all agents have about equal hypoglycemic potency.[1]

Mechanism of Action

The principal mode of action of sulfonylureas is to "solubilize" the release of stored insulin in response to a rising blood sugar level(2). For this reason they are only effective in relatively mild maturity-onset diabetics who still have substantial, albeit subnormal, insulin secretory capacity. Conversely, they are not very efficacious in more severe maturity-onset diabetics whose very high fasting blood sugar levels before treatment reflect severe loss of insulinogenic reserve; and they are completely useless in typical juvenile-onset diabetics who have no residual insulin secretory capacity.

Patients Most Likely to Respond

Diabetic patients most likely to respond to diet plus sulfonylurea therapy are the typical maturity-onset patients, who comprise 80% of all diabetics. Thus, they are older than age 40, and 80% of them are obese. In addition, evidences of substantial remaining insulin secretory capacity are that fasting plasma glucose levels are only moderately elevated, i.e. not over 300 mg per

---

[1] Until banned by the Food and Drug Administration in 1977, phenformin was also widely used, either singly or combined with a sulfonylurea agent, in dosages up to 100 mg daily.

100 ml; that the patients has never been in diabetic ketoacidosis; and, if on insulin therapy, that he is not taking more than 25 units daily. None of these are hard and fast rules, and it is worthwhile trying an essentially stable adult-onset diabetic on diet plus a sulfonylurea before resorting to diet and insulin therapy.

Diet as the Keystone of Therapy

When diet plus sulfonylurea therapy is used to obtain maximal blood sugar control, it is important to realize that the oral agents are potent enough to supplement dietary control but not to replace it. Since most maturity-onset diabetics are obese, the cardinal first step in management is to enforce a weight reduction regimen, which usually means a 1500 calorie intake for an obese male and 1200 calories for an obese female. Weight loss alone will substantially reduce fasting and postprandial plasma glucose levels in most adult-onset diabetics, and sometimes actually lowers them into the normal range. When plasma glucose levels remain elevated despite weight loss, sulfonylureas are then added for additional blood sugar control. On the other hand, if sulfonylureas are prescribed without dietary control, or if the patients regains lost weight while on a sulfonylurea, these agents are ineffective in controlling blood sugar levels.

Criteria of Blood Sugar Control

Most clinical studies in this country have used the criteria of the Joslin Clinic (3, 4), which measures "true glucose" (Somogyi-Nelson method) on venous whole blood. For good control, at least 70% of fasting or 3-hour postprandial blood sugars should not exceed 110 mg per 100 ml; for fair control, corresponding blood values should not exceed 130 mg per 100 ml; and poor control means respective values above the fair control limits. It should be noted that the UGDP study (discussed below) used these Joslin Clinic criteria of blood sugar control (5)[2].

EFFICACY OF SULFONYLUREAS

Sulfonylureas are effective when correctly used in properly selected patients. Since their main action is to enhance the patient's own insulin response to a rising blood sugar level, their efficacy clearly depends upon the adequacy of residual insulin secretory capacity. If the patient responds optimally to a combined diet and sulfonylurea program, and his blood sugar levels return to the normal range, as long as he stays on his diet and takes

[2] Plasma or serum values are 15% higher than whole blood values. Thus, corresponding fasting plasma levels by the Joslin criteria would be: good control, up to 125 mg per 100 ml; fair control, up to 150 mg per 100 ml; poor control, above 150 mg per 100 ml.

the oral agent regularly his blood sugar should remain controlled indefinitely. In other words, the effect of a sulfonylurea does not "wear off" with time, but continues to enhance insulin release as long as therapeutically effective levels of the drug are bathing the islets.

"Primary Failures"

Primary failures occur in maturity-onset diabetics when satisfactory blood sugar control is never achieved with diet plus maximal sulfonylurea therapy. In the early years of evaluating sulfonylureas in large series of adult-onset diabetics, when all degrees of hyperglycemic severity were being evaluated, the primary failure rates ranged from 16 to 36% (3, 4, 6). However, with better selection of patients as noted above, and especially with continued enforcement of dietary adherence, primary failure rate should be no more than 10–15%.

"Secondary Failures"

In evaluating the efficacy of sulfonylureas, true secondary failures must be distinguished from apparent ones. In a true secondary failure, the patient initially achieves optimal plasma sugar control on a reduction or maintenance diet and a maintenance dosage of sulfonylurea. Then, despite adherence to that program, including continued weight reduction if still obese, the plasma sugar rises and the maximal dosage of a single oral agent no longer controls it. Such true secondary failures of course occur, since in some patients the diabetes worsens spontaneously as insulin secretory capacity falls, but this is a relatively infrequent development in an originally responsive patient.

However, most so-called secondary failures are actually failures to stay on the diet. This is easily detected by weighing the patient at every visit, and finding that a gain of only a pound or so is accompanied by a rising blood sugar level. If the patient continues to gain weight, or even fails to lose weight again, hyperglycemia persists as expected. One can prove that the secondary failure is not real by hospitalizing the patient, enforcing a low calorie diet, and watching him regain responsiveness to the diet plus sulfonylurea program as he loses weight.

In this regard, in a recent review of oral hypoglycemic agents Shen & Bressler (7) emphasized the high rate of secondary failures on long-term sulfonylurea therapy. They noted that the Joslin Clinic had reported 25% secondary failures in 430 patients treated with tolbutamide for 6–9 years; and that, after the first 2 years, failure occurred in about 25% of the patients annually (4). In addition, during a 1–9 year follow-up at Boston City Hospital of 280 clinic patients who initially responded to tolbutamide or chlorpropamide therapy, and who actually needed the drug for blood sugar control, Singer & Hurwitz (6) found secondary failures in 29% of the tolbutamide patients and 26% of the chlorpropamide group. However, in neither of these long-term clinical studies was mention made of weight gain during the years of follow-up, as a probable contributing factor toward these secondary failures to maintain blood sugar control.

For this very reason, in reporting his series of 545 patients on diet plus chlorpropamide therapy for up to 12 years, Haunz (8) stressed that the single most important factor in long-term success was close follow-up every 1–2 months on an outpatient basis, with emphasis on continuous dietary control and lowest effective dosage of chlorpropamide. These were private patients followed personally by a meticulous clinician, in contrast to thousands of diabetics enrolled in large charity clinics and seen by a different staff physician at each visit during years of casual follow-up. Similar long-term good blood sugar control has been reported by other diabetologists working with private-patient populations (S. B. Beaser, C. R. Shuman, J. M. Moss, personal communications). In short, the type of patient care determines the difference between what can be done with diet plus sulfonylurea agents, and what too often is actually done with them.

The true effectiveness of sulfonylureas is also demonstrated by what happens when they are stopped in patients on maintenance therapy. Tsalikian et al (9) took 19 adult-onset diabetics off of oral agents; 12 of them had been on sulfonylurea alone and 7 others on sulfonylurea plus phenformin. These were ordinary clinic patients with the usual range from good to poor control on the cursory diet and oral agent program achievable in most large hospital clinics. On oral hypoglycemic therapy their mean fasting plasma sugar (FPS) was 227 mg per 100 ml, with a range of 95–555 mg per 100 ml; and average postprandial plasma sugar was 285 mg per 100 ml (range, 128–592 mg per 100 ml). Oral agents were then stopped abruptly, and circulating sugar levels were checked one week later and again three weeks later. Average fasting plasma glucose rose to 287 mg per 100 ml with a range of 143–678 mg per 100 ml and mean postprandial plasma sugar was 391 mg per 100 ml (range 216–822 mg per 100 ml). Off treatment, 16 of the 19 patients became symptomatic with weight loss, polyuria, polydipsia, weakness, blurred vision, or vaginitis; three of them required hospitalization for acute management of severe hyperglycemia. Thus, in this study oral agents were demonstrably effective in lowering blood sugar in adult-onset diabetics.

SAFETY OF SULFONYLUREAS

Other than minor side effects like skin rash, gastric intolerance, or diarrhea, the safety of sulfonylureas involves three important considerations; (*a*) water intoxication, (*b*) hypoglycemic coma, and (*c*) the validity of the conclusions of the University Group Diabetes Program (UGDP).

Water Intoxication

Symptomatic dilutional hyponatremia is a recognized complication of chlorpropamide therapy, but is not caused by the other sulfonylureas. Chlorpropamide causes water retention both by peripheral augmentation of the action of antidiuretic hormone and by increasing the release of this hormone (10). In contrast, acetohexamide and tolazamide have mildly diuretic actions. Thus, to avoid the potential hazard of chlorpropamide as a cause of water retention in elderly diabetics with cardiovascular, hepatic, or renal disease, one of the other sulfonylureas is preferable.

Hypoglycemic Coma

Severe refractory hypoglycemic coma has been produced by all sulfonylureas now used in the United States. In 1972 the author reviewed 473 cases of profound drug-induced hypoglycemic coma (11); and he recently updated a total of 778 cases reported through 1976 (12). In both reports the most common cause of severe hypoglycemia was sulfonylureas (220 cases in 1972 and 465 cases in 1976). In the earlier report, 25 of the 220 patients died, and 9 others sustained irreversible brain damage. Chlorpropamide caused more than half of the sulfonylurea hypoglycemias (120 patients), and almost half of the deaths or serious sequelae (16 patients). By 1976, 59 of the 465 reported cases of sulfonylurea-induced hypoglycemia had died or sustained permanent complications. Chlorpropamide was again the chief offender, having caused 219 hypoglycemic comas, including 26 deaths and/or sequelae; and the number of cases caused by the new agent, glyburide, had jumped (in Europe) from 7 to 78 cases in only four years (12).

Special clinical features of sulfonylurea-induced hypoglycemia are important to know. First of all, these agents should not be used in elderly patients who live alone or are poorly monitored in nursing homes; and they should also not be used in malnourished persons or those with hepatic or renal disease. Secondly, sulfonylurea hypoglycemia is extremely refractory to treatment. The patient must be kept on continuous 10% glucose infusion for days, even up to two weeks occasionally, until mounting hyperglycemia signals that the drug effect has worn off (11). Otherwise he relapses repeatedly into hypoglycemic coma. Thirdly, certain drugs potentiate the hypoglycemic action of co-administered sulfonylureas. The three main drugs causing this are sulfisoxazole (Gantrisin®), given for urinary tract infections; bishydroxycoumarin (Dicoumarol®) for anticoagulation; and phenylbutazone (Butazolidin®) for arthritis (11).

The University Group Diabetes Program Controversy

The nine-year-long controversy over the findings and interpretations of the UGDP study, and over the proposed federal actions based upon them, is best understood by considering separately the arguments first of the proponents of the UGDP study, and then of its opponents.

THE PRO-UGDP SIDE The UGDP was a long-term (10 years) prospective clinical study whose purpose was to determine whether blood sugar control prevented or ameliorated the vascular complications of maturity-onset diabetes (13). It was funded by the National Institutes of Health (NIH) and involved 12 university centers (clinics) and a coordinating center in Baltimore. The criteria for enrollment were diagnosis of diabetes within the previous year; estimated prognosis for life of at least five years; and diabetes of only moderate severity so that a patient's status would presumably not be jeopardized if left untreated (i.e. on placebo) for ten years. About 200 patients each were treated with placebo (PLBO), tolbutamide (TOLB), a standard dose of insulin (ISTD), a variable dose of insulin (IVAR), or phenformin (PHEN). Fixed dosages of TOLB, ISTD, and PHEN were used, and only IVAR was given as needed to control blood sugar levels. The first four treatment groups were started in the first seven clinics in January 1961; and in August 1962 phenformin was added to the seventh clinic and the last five treatment centers. All 1027 patients were enrolled in all clinics by February 1966.

By May 1968, it had become apparent that more total deaths and more cardiovascular (CV) deaths were occurring in the TOLB-treated patients. When subsequent statistical analysis confirmed that CV deaths in TOLB-treated patients were significantly greater than in PLBO-treated patients, TOLB administration was discontinued on October 7, 1969. The UGDP workers reported their findings on the 823 patients in the first four treatment groups (i.e. not PHEN) to the American Diabetes Association (ADA) in June 1970 (Table 2). The Food and Drug Administration (FDA) had already announced its support of the UGDP indictment of TOLB as a cardiotoxic agent. In October 1970 both the ADA and the American Medical Association officially supported the UGDP findings; and the FDA proposed a labeling change in the package insert of all sulfonylurea agents, based solely upon the UGDP findings. The package insert warned clinicians about the potential cardiovascular toxicity of tolbutamide and all sulfonylureas, and recommended restricting their use to patients who required more than dietary control alone but in whom insulin was "impractical or unacceptable." In November 1970 the UGDP data on the tolbutamide findings were published (5).

In January 1971, phenformin was also discontinued; subsequently a preliminary report (14) announced similarly excessive CV deaths in PHEN-treated patients (26 CV deaths in PHEN patients compared to 11 CV deaths in control patients). Thus, two entirely different oral hypoglycemic agents stood indicted as cardiotoxic agents.

Table 2 Cumulative deaths in UGDP treatment groups

Date (reference)	PLBO[a] (205)[b]	TOLB (204)	PHEN (204)	ISTD (210)	IVAR (204)
Cardiovascular deaths					
October 7, 1969 (5)	10	26	—[c]	13	15[d]
March 7, 1973 (16)	26	35	31	22	21
December 31, 1974 (18)	29	—	—	27	29
Total deaths					
October 7, 1969 (5)	21	30	—	20	21[d]
March 7, 1973 (16)	45	45	41	33	32
December 31, 1974 (18)	54	—	—	48	49

[a] Treatment groups: placebo, tolbutamide, phenformin, standard insulin, variable insulin.
[b] Number of patients per treatment group.
[c] No information.
[d] Originally reported as 12 cardiovascular and 18 total deaths (5); corrected by FDA audit (19).

In June 1972, the Biometric Committee was appointed by the NIH to review the UGDP data and interpretations. This distinguished panel carefully scrutinized the original mortality findings of the 823 patients in the first four treatment groups, reviewed the statistical analyses of those data. and certified their validity. In their published report (15), the one new finding was that the cardiovascular toxicity of TOLB was only significantly excessive in women over 53 years of age.

The complete report on PHEN-treated patients appeared in February 1975 (16). This paper contained total and cardiovascular mortality data in all five treatment groups, as of March 7, 1973 (Table 2). The new findings indicated that, between October 1969 and March 1973, deaths from all causes in the PLBO-treated groups had caught up with those in the TOLB patients (45 each); and that there had been 16 CV deaths in the PLBO group compared to only 9 more in the TOLB patients. This report also disclosed that among PHEN-treated patients there had been three cases of lactic acidosis, one of which was fatal.

In July 1977, the FDA virtually banned the continued clinical use of phenformin, citing it as an "imminent hazard" to health (17). This action was based on several reports of excessive deaths from lactic acidosis in patients treated with phenformin, including the single such fatality in the UGDP report. However, the UGDP allegation of cardiotoxicity was not invoked as a reason for removing phenformin from use.

Subsequently, in July 1978 the UGDP workers reported that even insulin therapy was ineffective in managing adult-onset diabetes (18). Their statistical analysis indicated that even in IVAR patients, in whom insulin dosage was presumably given as required to control blood sugar levels, cardiovascular mortality was the same as in the PLBO group (29 CV deaths in each group). Thus, the overall conclusions of the entire UGDP study were that no treatment except dietary therapy is effective for blood sugar control in adult-onset diabetes.

In December 1978 an audit committee appointed by the FDA confirmed the conclusions of the UGDP study on oral hypoglycemic drugs (19). Its specific mission was to determine if the baseline and follow-up data submitted from the 12 study centers had been correctly logged into the computer in Baltimore. It found that, although "certain errors and discrepancies" did exist between the submitted and the recorded data, they were not of sufficient magnitude to negate the overall significance of the UGDP conclusions.

THE ANTI-UGDP SIDE In November 1970 about 40 (since grown to 250) vociferous opponents of the UGDP study, and the impending FDA action based upon its findings, formed the Committee for the Care of the Diabetic (CCD). It consisted of clinical and academic diabetologists and various consultants who believed that both the design and execution of the study were so seriously flawed that valid interpretations were impossible, and that a concerted effort should be made to prevent federal restrictions against the use of sulfonylureas. In September 1972 the author summarized more than 40 criticisms of the UGDP study (20), all based on published UGDP data. Several of the main criticisms were: (*a*) Patient populations in the 12 treatment centers were not homogeneous (i.e. different centers enrolled patients from charity clinics, mass screening programs, or private practice). Thus, patients differed greatly in baseline health status, especially regarding antecedent cardiovascular disease. (*b*) Baseline risk factors were also unevenly distributed among the five treatment groups. In particular, 9 of 10 risk factors occurred more often in the TOLB group than in the PLBO patients (21, 22). (*c*) The clinical prognosis of many patients was clearly less than the required five years. Forty-six percent of these adult-onset diabetics were 55 to 79 years old on admission, and 46% of them had known cardiovascular disease. (*d*) The use of fixed amounts of TOLB (1.5 g daily), PHEN (100 mg daily), and ISTD (10–16 units daily, based on body surface area) resulted in an unknown number of patients per group being overtreated or undertreated, in terms of blood sugar control, during years of follow-up. (*e*) Deaths were assigned to their original treatment groups, both in patients who died on a different treatment and in those who dropped out. Thus, two cardiovascular deaths attributed to the TOLB group occurred after 5½ and 6½ years of follow-up, in patients who had taken the drug for only 3 months and 21 months, respectively. (*f*) Just two centers contributed half of both the total deaths (15 of 30) and the CV deaths (13 of 26) in TOLB-treated patients. Patients in these two clinics also had the highest number of baseline risk factors.

The most complete analyses of both the clinical and statistical fallacies of the UGDP study were published in a quartet of papers by Feinstein (23–26). In addition, the medicolegal ramifications of the UGDP controversy have been lucidly summarized by Kolata (27).

Finally, three especially devastating analyses of the UGDP data were recently published by Kilo, Williamson, and their associates (28–30). As a co-principal investigator at the St. Louis center of the UGDP study, Kilo was able to obtain almost complete computer records of most of the study patients who had died. Their independent analysis yielded the following new information. (*a*) They confirmed in more detail the previous claim by others (21, 31) that it was the spuriously low number of CV deaths in PLBO patients (only 10 deaths) during the "tolbutamide phase" of the study (through December 1969) which established the statistical significance of the 26 CV deaths in the TOLB group. (*b*) When patients who had dropped out or changed medication were excluded from analysis, virtually all of the excessive CV mortality in the TOLB group occurred in patients whose fasting blood sugar (FBS) values chronically exceeded 200 mg per 100 ml; i.e. who were in "poor control" by the UGDP's own definition of the latter as FBS values above 130 mg per 100 ml (5). Conversely, CV mortality in TOLB patients with chronic FBS below 200 mg per 100 ml was the same as in both insulin-treated groups. (*c*) When PLBO and IVAR patients with comparable baseline risk factors were matched, there were 12 CV deaths in the PLBO group compared to only 3 CV deaths in the IVAR patients; this indicates that properly administered insulin therapy is indeed efficacious in reducing CV mortality in relatively mild diabetics (28–30). Kilo and Williamson concluded that, when other cardiovascular risk factors were eliminated, the UGDP data actually showed beneficial effects of blood sugar control on prognosis for life in both the TOLB and the IVAR groups, compared to the PLBO controls.

CURRENT STATUS In November 1978 the FDA issued its "final recommendation" that the long debated labeling change regarding inclusion of the possible cardiotoxicity of sulfonylureas be put into effect (32). The directive specifically stated that, although the FDA audit of the UGDP data had uncovered various discrepancies between data recorded in computer printouts and the published data, the latter were not serious enough to invalidate the significance of the excessive cardiovascular deaths in TOLB versus PLBO patients in October 1969. This edict led the American Diabetes Association to reconsider its former support of the UGDP findings. A committee of the ADA reviewed all the pros and cons of the UGDP controversy, and in January 1979 officially recommended that the FDA withhold any decision regarding the continued use of sulfonylureas until it carefully considers all of the UGDP data, both before and after October 1969; and that select committees from both the FDA and ADA be assigned to resolve the UGDP controversy once and for all (33). The latter cooperative effort is now under way.

SUMMARY

(*a*) Properly used, sulfonylureas are both indefinitely effective and entirely safe. (*b*) Diet is the keystone of treatment—especially now that combined sulfonylurea and phenformin cannot be used. (*c*) True "secondary failures" are rare. (*d*) The only real risk from sulfonylureas is drug-induced hypoglycemia, which is especially prevalent with chlorpropamide. (*e*) The conclusions of the UGDP study are becoming increasingly untenable.

ACKNOWLEDGMENT

The author would like to thank Ms. Ann Burns for her help in preparing the manuscript.

Literature Cited

1. Hamwi, G. J., Seidensticker, J. F. 1967. Therapy: Oral hypoglycemic agents. In *Diabetes Mellitus: Diagnosis and Treatment, Vol. 2,* ed. G. J. Hamwi, T. S. Danowski, pp. 109–15. New York: American Diabetes Association. 250 pp.
2. Seltzer, H. S. 1962. Quantitative effects of glucose, sulfonylureas, salicylate, and indole-3-acetic acid on the secretion of insulin activity into pancreatic venous blood. *J. Clin. Invest.* 41:289–300
3. Camerini-Davalos, R. A., Marble, A. 1962. Incidence and causes of secondary failure in treatment with tolbutamide. Experience with 2,500 patients treated up to five years. *J. Am. Med. Assoc.* 181:89–92
4. Balodimos, M. C., Camerini-Davalos, R. A., Marble A. 1966. Nine years' experience with tolbutamide in the treatment of diabetes. *Metabolism* 15: 957–70
5. Klimt, C. R., Knatterud, G. L., Meinert, C. L., Prout, T. E. 1970. The University Group Diabetes Program: A study of the effects of hypoglycemic agents on vascular complications of patients with adult-onset diabetes. II. Mortality results. *Diabetes* 19 (Suppl. 2). pp. 789–830
6. Singer, D. L., Hurwitz D. 1967. Long-term experience with sulfonylureas and placebo. *N. Engl. J. Med.* 277:450–56
7. Shen, S.-W., Bressler, R. 1977. Clinical pharmacology of oral antidiabetic agents (second of two parts). *N. Engl. J. Med.* 296:787–93
8. Haunz, E. A. 1973. Chlorpropamide hypoglycaemia. *Lancet* 2:46–47
9. Tsalikian, E., Dunphy, T. W., Bohannon, N. V., Lorenzi, M., Gerich, J. E., Forsham, P. H., Kane, J. P., Karam, J. H. 1977. The effect of chronic oral antidiabetic therapy on insulin and glucagon responses to a meal. *Diabetes* 26: 314–21
10. Moses, A. M., Howanitz, J., Miller, M. 1973. Diuretic action of three sulfonylurea drugs. *Ann. Intern. Med.* 78: 541–44
11. Seltzer, H. S. 1972. Drug-induced hypoglycemia. A review based on 473 cases. *Diabetes* 21:955–66
12. Seltzer, H. S. 1979. Severe drug-induced hypoglycemia: A review. *Compr. Ther.* 5:21–29
13. Klimt, C. R., Knatterud, G. L., Meinert, C. L., Prout, T. E. 1970. University Group Diabetes Program: A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. I. Design, methods and baseline results. *Diabetes* 19 (Suppl. 2), pp. 747–83
14. Knatterud, G. L., Meinert, C. L., Klimt, C. R., Osborne, R. K., Martin, D. B. 1971. Effects of hypoglycemic

agents on vascular complications in patients with adult-onset diabetes. IV. A preliminary report on phenformin results. *J. Am. Med. Assoc.* 217:777–84

15. Report of the committee for the assessment of biometric aspects of controlled trials of hypoglycemic agents. 1975. *J. Am. Med. Assoc.* 231:583–608
16. Knatterud, G. L., Klimt, C. R., Osborne, R. K., Meinert, C. L., Martin, D. B., Hawkins, B. S. 1975. The University Group Diabetes Program: A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. V. Evaluation of phenformin therapy. *Diabetes* 24 (Suppl. 1), pp. 65–184
17. Califano, J. A. Jr. 1977. Order of the Secretary suspending approval. In re: New drug applications for phenformin. (July 25)
18. Knatterud, G. L., Klimt, C. R., Levin, M. E., Jacobson, M. E., Goldner, M. G. 1978. Effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. VII. Mortality and selected nonfatal events with insulin treatment. *J. Am. Med. Assoc.* 240: 37–42
19. Audit confirms conclusions of UGDP study on oral diabetes drugs. 1978. *FDA Drug Bull.* 8:34–36 (Dec. 1978–Jan. 1979)
20. Seltzer, H. S. 1972. A summary of criticisms of the findings and conclusions of the University Group Diabetes Program (UGDP). *Diabetes* 21:976–79
21. Schor, S. 1971. The University Group Diabetes Program: A statistician looks at the mortality results. *J. Am. Med. Assoc.* 217:1671–75
22. Bradley, R. F., Dolger, H., Forsham, P. H., Seltzer, H. 1975. "Settling the UGDP controversy"? *J. Am. Med. Assoc.* 232: 813–17
23. Feinstein, A. R. 1971. Commentary: Clinical biostatistics VIII. An analytic appraisal of the University Group Diabetes Program (UGDP) study. *Clin. Pharmacol. Ther.* 12:167–91
24. Feinstein, A. R. 1976. Clinical biostatistics XXXV. The persistent clinical failures and fallacies of the UGDP study. *Clin. Pharmacol. Ther.* 19:78–93
25. Feinstein, A. R. 1976. Clinical biostatistics XXXVI. The persistent biometric problems of the UGDP study. *Clin. Pharmacol. Ther.* 19:472–85
26. Feinstein, A. R. 1979. How good is the statistical evidence against oral hypoglycemic agents? *Adv. Intern. Med.* 24: 71–95
27. Kolata, G. B. 1979. Controversy over study of diabetes drugs continues for nearly a decade. *Science* 203:986–90
28. Kilo, C., Williamson, J. R., Choi, S. C., Miller, J. P. 1979. Letter to the editor: Insulin treatment and diabetic vascular complications. *J. Am. Med. Assoc.* 241: 26–27
29. Kilo, C., Williamson, J. R., Choi, S. C., Miller, J. P. 1979. Refuting the UGDP conclusion that insulin treatment does not prevent vascular complications in diabetes. In *Proc. 4th Int. Symp. on Early Diabetes* ed. R. A. Camerini-Davalos. New York: Academic. In press
30. Kilo, C., Miller, J. P., Williamson, J. R. 1979. The Achilles heel of the University Group Diabetes Program. *J. Am. Med. Assoc.* In press
31. O'Sullivan, J. B., D'Agostino, R. B. 1975. Decisive factors in the tolbutamide controversy. *J. Am. Med. Assoc.* 232:825–29
32. Kennedy, D. 1978. Food and Drug Administration (FDA) analysis of UGDP study. *Fed. Reg.* 43 (220):52732–34 (Nov. 14)
33. Whitehouse, F. W., Arky, R. A., Bell, D. I., Lawrence, P. A., Freinkel, N. 1979. Policy statement. The UGDP controversy. *Diabetes* 28:168–70

Ann. Rev. Med. 1980, 31:273–88
Copyright © 1980 by Annual Reviews Inc. All rights reserved

IDIOPATHIC MEMBRANOPROLIFERATIVE GLOMERULONEPHRITIS

❖7360

Youngki Kim, M.D. and Alfred F. Michael, M.D.

Department of Pediatrics, University of Minnesota Medical School,
Minneapolis, Minnesota 55455

INTRODUCTION

Idiopathic membranoproliferative glomerulonephritis (MPGN) is one identifiable form of chronic nephritis characterized by distinctive glomerular morphology. However, it is not known whether MPGN is a single pathogenetic entity. It was originally recognized as one type of chronic glomerulonephritis with persistent hypocomplementemia by West et al (1) and Gotoff et al (2). Subsequently it became apparent that hypocomplementemia, though common, was not invariably present in every patient.

Although similar in their clinical manifestations, two main types of MPGN have been recognized based upon differences in ultrastructural morphology: type I MPGN with subendothelial deposits and type II MPGN (dense deposit disease) with unique intramembranous dense deposits. A third type of MPGN characterized by the presence of subepithelial and subendothelial deposits has also been described (3, 4). This review discusses types I and II MPGN.

PATHOLOGY

The characteristic renal histology of types I and II MPGN has been previously described in detail (1, 5–11).

Type I MPGN

Light microscopy reveals diffuse enlargement of glomerular tufts often with lobulation (Figure 1*a*). Mesangial cellularity and matrix are increased. Glomerular capillary walls look thickened because of the interposition of mesangial matrix between glomerular basement membrane (GBM) and

0066-4219/80/0401-0273$01.00

endothelium. On periodic acid-Schiff or silver staining, however, the GBM is of normal thickness. The normal GBM together with the interposed mesangial matrix gives appearance of splitting of glomerular capillary walls (double contour). Neutrophils may be seen in increased number in glomeruli. Epithelial crescents may also be present, and are usually indicative of a poor prognosis. Changes in tubules and interstitium are not remarkable. Electron microscopy shows an increase in mesangial cellularity and matrix, swelling of mesangial cells, and interposition of mesangial matrix (Figure 1*c*). The GBM is of normal thickness, and fine granular electron dense deposits are present along the subendothelial side of the GBM, and within mesangium. Hump-like subepithelial deposits may occasionally be present. Immunofluorescence microscopy reveals granular deposition of C3 and properdin in the mesangium, and along the peripheral GBM in all patients (Figure 1*b*, Table 1). Immunoglobulins and other complement components are also found in this location, but in lesser frequency and intensity.

Type II MPGN

Light microscopy reveals diffuse hypertrophy of glomerular tufts, and increase in mesangial cellularity and matrix frequently to a lesser degree than seen in type I (Figure 1*d*). Neutrophils can also be seen in glomeruli, and epithelial crescents are more frequently present than in type I MPGN. There is marked irregular ribbon-like thickening of glomerular capillary wall due to the presence of dense deposit material (DDM). These dense deposits are refractile when stained with hematoxylin-eosin, and are stained intensely with periodic acid-Schiff; dense deposits also may be present in Bowman's capsule, tubular basement membrane, and mesangium. Electron microscopy reveals homogeneous and strongly electron-dense material in the same distribution as seen by light microscopy (Figure 1*f*). In addition, subepithelial deposits (humps) are frequently observed; subendothelial deposits are scanty or absent. By immunofluorescence, a specific finding in all patients with type II, but not type I MPGN, is the presence of C3 along the margin, but not within the dense deposit material, giving a double linear appearance to the glomerular capillary (railroad tracks); C3 is also present within the mesangium outlining circular dense deposit material (mesangial rings) (Figure 1*e*, Table 1). These C3 railroad tracks and mesangial rings also contain properdin and C4, but never any immunoglobulins. Therefore, by electron and immunofluorescence microscopy, the DDM in type II MPGN would appear to be different than the granular immune deposits observed in type I MPGN and other forms of glomerulonephritis. However, granular immune deposits of immunoproteins similar to those observed in type I MPGN may also be seen in type II disease but are quantitatively less. The nature of the dense deposit material is not yet known.

CLINICAL MANIFESTATIONS

Type I MPGN

The disease occurs in older children, usually over the age of 8 years, and adults but has previously been reported as early as 22 months of age (7, 10, 12, 13). The onset of the disease may manifest in several ways (7, 10, 12,

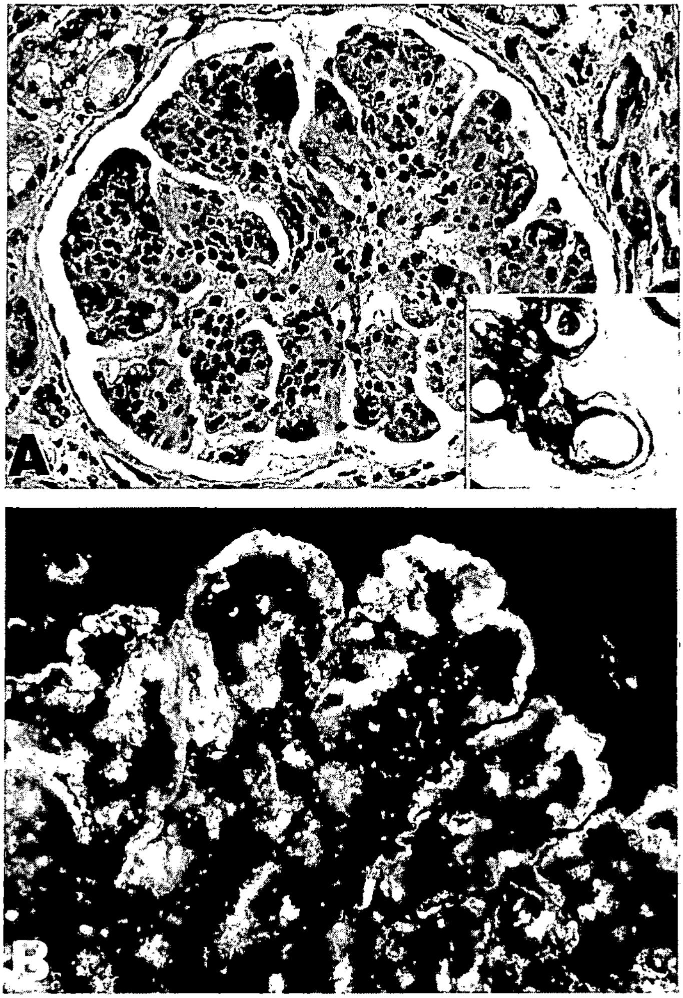


Figure 1a,b (*a*) Light microscopy showing an enlarged glomerulus with increase in mesangial cellularity and matrix, and splitting of glomerular capillary walls (insert) in type I MPGN (H & E, X250). (*b*) Immunofluorescence microscopy; stained with fluorescein-conjugated anti-C3 showing granular deposition of C3 along the GBM and in the mesangium in type I MPGN (X610).

14–16): nephrotic syndrome, gross hematuria, or asymptomatic proteinuria and microscopic hematuria (Table 2). Unlike type II MPGN, acute nephritic episodes (oliguria, edema, hematuria, hypertension, and renal insufficiency) are rather uncommon (16). A history of infections preceding the onset, especially those of upper respiratory tract can be obtained in a third of the patients. Although elevated streptococcal antibody titers may be seen

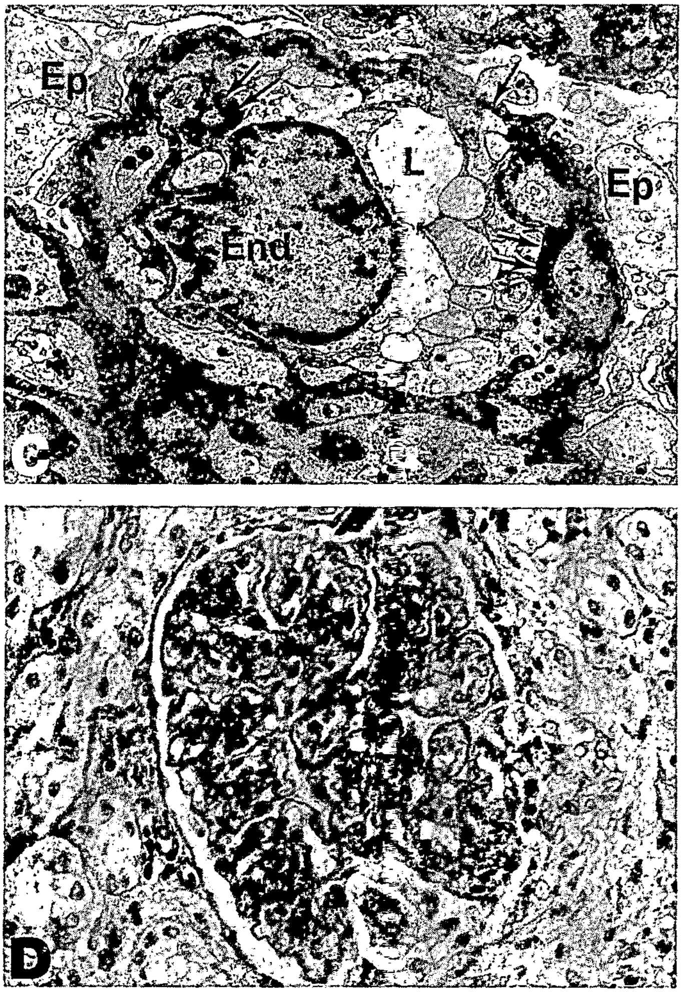


Figure 1c,d (*c*) Electron micrograph of a portion of glomerulus showing normal GBM (*single arrow*) with mesangial interposition (*double arrows*) in type I MPGN: L = capillary lumen, Ep = epithelial cell, End = endothelial cell (X6000). (*d*) Light microscopy showing thickening of GBM due to the presence of dense deposit material in type II MPGN. Note that the mesangial changes are less than those in type I MPGN (periodic acid-Schiff, X325).

occasionally, there is no evidence that this finding is more frequently seen than in the general population or that the disease itself is a consequence of a preceding infection. In some circumstances, however, poststreptococcal glomerulonephritis may be difficult to distinguish from MPGN on clinical and morphological grounds. At some time during the course of the disease,

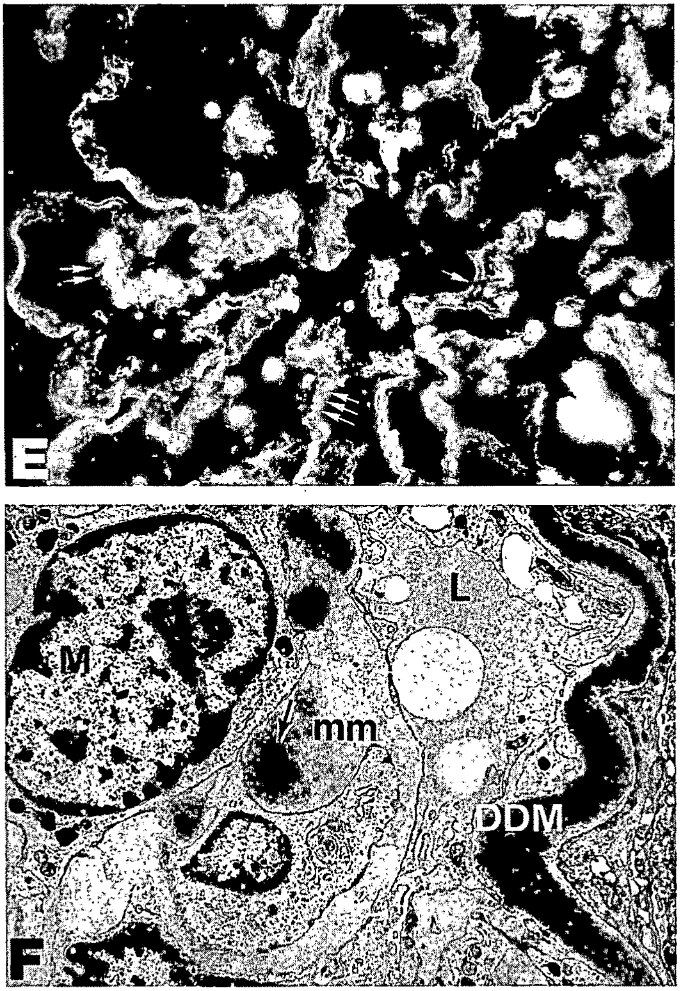


Figure 1e,f (*e*) Immunofluorescence microscopy showing C3 railroad tracks (*single arrow*) and mesangial rings (*double arrows*) in type II MPGN. Granular deposits of C3 along the GBM can also be seen (*triple arrows*) (X670). (*f*) Electron micrograph showing DDM occupying the GBM in type II MPGN. Note also the presence of DDM in mesangium (*single arrow*): DDM = dense deposit material, L = capillary lumen, mm = mesangial matrix, M = mesangial cell nucleus (X6000).

Table 1 Immunohistologic changes in glomeruli in MPGN

	Type I		Type II	
	Granular deposit	Railroad tracks and/or mesangial rings	Granular deposit	Railroad tracks and/or mesangial rings
IgG	8/9[a]	0/9	4/8	0/8
IgM	7/9	0/9	8/8	0/8
IgA	2/9	0/9	1/8	0/8
Clq	3/9	0/9	0/8	0/8
C4	9/9	0/9	8/8	4/8
C2	7/9	0/9	2/8	0/8
C3	9/9	0/9	8/8	8/8
Properdin	9/9	0/9	8/8	5/8
Factor B	0/9	0/9	0/8	0/8
Fibrin	8/9	0/9	5/8	0/8
Albumin	0/9	0/9	0/8	0/8

[a] Denominator indicates the number of patients studied and numerator the number of patients with positive findings.

almost all patients develop hematuria (macroscopic or microscopic) and varying degrees of proteinuria with approximately 80% of patients demonstrating features of the nephrotic syndrome, which is usually mild. Hypertension is present at onset in approximately a third of the patients and is commonly observed as the disease progresses.

In general, the clinical course of the disease is quite variable (7, 12, 15, 17, 18). While some patients maintain persistent nephrotic syndrome, others have intermittent nephrotic or nephritic episodes with abnormal urinary findings between the episodes. Occasionally, patients with an acute nephritic onset may become completely asymptomatic with normal renal function and urinary findings (9, 18, 19). This so-called silent phase of the disease may last for several years despite the persistent morphological abnormalities of the kidney tissues. In general, the prognosis is worse in patients with nephrotic syndrome, persistent hypertension, or gross hematuria (7, 12, 20, 21). In a follow-up study of 84 patients for a mean duration of 6 years, Habib et al (10) reported that 25% were either dead from renal insufficiency or were on chronic hemodialysis, 11% were in chronic renal failure, 29% had nephrotic syndrome, 33% had only proteinuria, and 5% appeared to be in spontaneous though transient remission. Repeat kidney biopsies in the latter group rarely showed significant morphological changes regardless of the interval. We have observed complete clinical remission in 5 of 63 patients with type I MPGN (Y. Kim, A. F. Michael, unpublished observations).

Table 2 Comparison of typical type I and II MPGN[a]

Finding	Type I	Type II
Clinical		
Acute nephritic episode	uncommon	common
Nephrotic syndrome	common	common
Hematuria	common	common
Hypertension	common	common
Recurrence in transplanted kidneys	occasional	usual
Morphological		
Subendothelial deposit	present	absent
Intramembranous dense deposit	absent	present
C3 mesangial rings and railroad tracks	absent	present
Complement		
C3	normal or reduced	reduced
C1 and C4	reduced	normal
C3NeF	absent	present

[a] Except for the morphological findings described, significant overlap between the two types may exist.

Type II MPGN

Clinical manifestations in both types I and II MPGN are quite similar, making it extremely difficult to distinguish one from the other on clinical grounds alone (8, 10, 12, 13, 16, 22, 23). However, there are several differences (Table 2) (10, 12, 16): (*a*) age of the onset tends to be younger in type II (usually between 8 and 25 years of age) than in type I MPGN, (*b*) asymptomatic proteinuria and hematuria at the onset of the disease are not as common in type II as in type I MPGN, (*c*) persistent nephrotic syndrome, macroscopic hematuria, and acute nephritic episodes are more common in type II than in type I MPGN, and (*d*) prognosis is worse in type II than in type I MPGN because of the higher incidence of acute nephritic episodes and the more frequent presence of epithelial crescents in glomeruli. In a study of the actuarial survival (10), a 50% mortality occurred between the ninth and tenth year of the disease for type II MPGN and between the seventeenth and eighteenth year for type I MPGN.

Another interesting feature is the frequent association of type II MPGN with partial lipodystrophy (22, 24–26). Patients with partial lipodystrophy can be divided into three groups: (*a*) no complement abnormalities, (*b*) hypocomplementemia with high serum concentrations of C3 nephritic factor (C3NeF), but without renal disease, and (*c*) hypocomplementemia and renal disease. The associated renal disease has always been type II MPGN except in two patients who had the type I variety (26, 27).

COMPLEMENT ABNORMALITIES

The complement system may be divided into classical and alternative pathways (Figure 2) (28). The classical pathway is composed of 11 serum proteins (C1–C9), and activated by immune complexes or aggregates of IgG and IgM. There are at least four serum proteins in the alternative pathway, which includes Factor B and properdin. The alternative pathway bypasses the early components of the classical pathway (C1, C4, C2), enters the complement cascade at the level of C3 with subsequent activation of C5–C9, and is activated by IgA aggregates or polysaccharide such as zymosan and endotoxin. Activation of either pathway results in the breakdown of C3 into two fragments, C3a and C3b.

In type I MPGN, levels of serum complement tend to fluctuate; it is not uncommon for a patient to have low complement values at one time, then later normal values, or vice versa. At diagnosis, decreased serum concentration of C3 is found in approximately a third to a half of the patients with type I MPGN (7, 11, 29, 30). However, serial determinations usually reveal hypocomplementemia sometime during the course of the disease in the majority of patients. In addition, serum concentrations of early components of the classical pathway (C1q, C4) and components of the alternative pathway (Factor B, properdin) are also frequently decreased.

In type II MPGN, serum concentration of C3 remains persistently low in almost all patients (7, 10, 11, 29, 30). Occasional patients, however, may have fluctuation of C3 values or even remain normocomplementemic throughout the course of the disease. In contrast to type I MPGN, serum

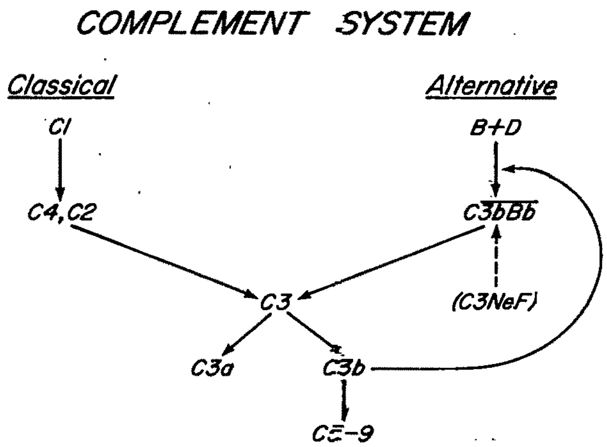


Figure 2 Simplified schema of the classical and alternative complement pathways. C3NeF is an autoantibody with specificity for C3b$\overline{\text{Bb}}$.

concentrations of the early components of the classical pathway (C1q, C4) are usually normal, although the components of the alternative pathway (Factor B, properdin) are decreased. This profile suggests that complement activation in type I MPGN occurs through the classical and alternative pathways, whereas in type II MPGN it occurs primarily through the alternative pathway.

Another abnormality of the complement system in patients with MPGN is the presence of C3 nephritic factor (C3NeF) (31, 32). C3NeF is an autoantibody of IgG class directed against the alternative pathway C3 convertase ($C\overline{3bBb}$), and activates the alternative pathway by stabilizing $C\overline{3bBb}$ (33–37). While C3NeF is present in high concentration in almost all patients with type II MPGN, it is found in lower concentrations in less than a third of the patients with type I MPGN (11, 29, 38, 39). The origin of C3NeF is unknown, but it is not derived from the kidney since it persists in circulation after bilateral nephrectomy (40). C3NeF itself is not nephritogenic as evidenced by the absence of renal disease in some patients with partial lipodystrophy and hypocomplementemia (41). Further, an infant born to a mother with MPGN and high concentration of C3NeF in her serum was found to be hypocomplementemic during the newborn period as a result of placental transfer of maternal C3NeF, but did not develop any overt renal disease (42).

The pathogenetic event leading to hypocomplementemia is also confusing. Metabolic studies utilizing radiolabeled C3 showed decreased synthetic rate of C3 in these patients (43–45). Furthermore, Colten et al demonstrated decreased synthesis of C3 in vitro by utilizing short-term culture of fresh liver biopsy material from patients with MPGN (46). On the other hand, evidence for in vivo complement activation has been deduced by the demonstration of (*a*) C3NeF in the serum, (*b*) cryoglobulins and circulating immune complexes in both types of MPGN (47, 48), and (*c*) circulating complement breakdown products in plasma (44, 49–51). The data are consistent with the view that hypocomplementemia reflects both increased catabolism and decreased synthesis as shown by Charlesworth et al (52). Despite the marked abnormalities of the complement system, there is no correlation between the serum levels of complement or C3NeF, and the clinical activity or progression of MPGN (9, 53–55).

RECURRENCES IN TRANSPLANTED KIDNEYS

The literature reports numerous recurrences of type I MPGN in transplanted kidneys (18, 56–59). It is, however, extremely difficult to determine the true incidence of recurrence because of the similarities of clinical and histological findings between type I MPGN and chronic graft rejection (57, 60–63). The relatively high frequency (approximately 50%) of membranoproliferative disease in transplanted kidneys reported by McLean et al (59) stands in contrast with the lower incidence in another report (approximately 18%) (61). These differences may be related to variables of the patient population and the likely heterogeneity of the syndrome, differing criteria for the histological diagnosis of recurrent disease, and an inadequate appreciation of morphological changes that occur in non-MPGN transplant population. In a study of glomerular morphology in transplanted kidneys, Mathew et al (60) noted that of nine patients with histological features compatible with type I MPGN, only two had type I MPGN as the original disease. This finding suggests that the true incidence of posttransplant recurrences for type I MPGN may be lower than generally thought. Furthermore, typical morphology of type I MPGN may occur in transplanted kidneys of patients who initially do not have type I disease in association with vesicoureteral reflux (64). A number of possibilities exist to explain these observations: the development or activation of infections in immunosuppressed transplant patients, leading to glomerulonephritis independent of the initial disease; the occurrence of homograft rejection; and the lack of specificity of mesangial proliferation and capillary wall thickening as indicative of a single disease process.

Another unsettled issue is whether pre- and posttransplant levels of serum complement have any significance in predicting the frequency of recurrence. While Berthoux et al (58) found an increased incidence of recurrence in patients with pre- and posttransplant hypocomplementemia, McLean et al (59) found no differences. Although recurrence may be associated with proteinuria, nephrotic syndrome, hypertension, or renal insufficiency, nephrectomy is infrequently indicated unless there is superimposed chronic graft rejection. In contrast to type I, type II MPGN recurs almost invariably, and dense deposit material can be seen in transplanted kidney as early as two weeks after transplant (65–68). Since the histological changes of type II MPGN have never been associated with any other disease, including chronic graft rejection, it is easier to recognize recurrence in the transplanted kidneys. It is of interest that histological changes of these kidneys tend to be benign despite the presence of dense deposit material in the basement membrane and mesangium (67). Symptoms associated with recurrence are also frequently mild or even absent (66–68). In their review of the world literature, Cameron & Turner (65) found nephrotic syndrome in only 5 of 22 reported cases, which is in sharp contrast to frequency in patients prior to transplant. However, we have seen one patient who developed severe recurrence of disease in two consecutively transplanted kidneys, which necessitated nephrectomy (Y. Kim, A. F. Michael, unpublished observation). Although reduction of C3NeF levels and increase in serum complement are frequently observed after transplantation, probably as a result of steroid therapy (55), pre- and posttransplant complement profiles have not been of value in predicting or recognizing recurrences (66, 69). Since recurrences infrequently cause graft failure in patients with type I and II MPGN, patients should not be denied renal transplant on the grounds of recurrence of the original disease.

PATHOGENESIS

In type I MPGN, available evidence suggests an immune complex renal injury: (*a*) immunofluorescence findings of granular deposition of immunoproteins in the kidney, (*b*) presence of cryoglobulins and C1q-binding immune complexes in serum, and (*c*) occurrence of glomerulonephritis morphologically similar to type I MPGN in immune complex diseases such as systemic lupus erythematosus and chronic bacteremia (5, 7). In addition, Kuriyama was able to produce similar renal lesions experimentally in rabbits by chronic injections of egg albumin (70). In this chronic serum sickness model, the rabbits that produced precipitating antibody of high avidity developed MPGN, while those with nonprecipitating antibody of low avidity developed membranous nephropathy. The clinical, morphological, and serological similarities between type I MPGN and glomerulonephritis associated with certain defined infections (e.g. shunt and hepatitis B nephritis) suggest a common pathogenesis resulting from chronic exposure to an infectious antigen, possibly viral. In this context, a number of different antigens might be responsible for the syndrome.

Immune complexes also play a role in type II MPGN, as evidenced by the immunofluorescence findings of granular deposition of immunoproteins and the demonstration of cryoglobulins and immune complexes in sera of these patients. However, the nature of dense deposit material is far from clear. Galle & Mahieu (71) showed that the biochemical composition of the GBM containing DDM is similar to that of normal GBM, except for increased sialic acid and decreased cystine content. Based on this finding, they suggested that DDM may represent accumulation of biochemically modified glycoprotein of normal basement membrane. The recurrence of type II MPGN in the transplanted kidney is compelling evidence that a systemic factor plays a crucial role. It is equally possible that DDM results from the accumulation of serum protein(s) or other materials in the native or denatured state. However, antiserum against human serum proteins including immunoglobulins and complements do not react with DDM by immunofluorescence technique. In addition, we have not been able to demonstrate reactivity of DDM with antisera to GBM, collagen, or antiserum raised to GBM containing DDM (11; Y. Kim, J. I. Scheinman, A. F. Michael, unpublished observations). It has been shown recently that the fluorescent dye thioflavin T stained DDM intensely (72). However, DDM failed to react with antisera against amyloid proteins (anti-SAA and anti-SAP), which suggests that the reactions of thioflavin T with amyloid and DDM are different (Y. Kim, A. F. Michael, unpublished observation). Immune complexes may be the initial pathogenetic events in both type I and II MPGN, and dense deposit transformation may be a consequence of a unique host response to this immunologic injury. On the other hand, DDM may represent a host protein that has lost its antigenic determinants and is no longer recognizable by immunohistologic techniques.

TREATMENT

In most of the published studies, the results of treatment with corticosteroid and/or cytotoxic drugs such as cyclophosphamide and azathioprine have been disappointing (7, 9, 12, 17).

Kincaid-Smith (73) reported a favorable response to relatively long-term (up to 3 years) treatment with combined cyclophosphamide, dipyridamole, and anticoagulants in adults with types I and II MPGN. All of these patients had nephrotic syndrome, hypertension, and impaired renal function; the combined regimen improved renal function, clinical course, histology, and survival rates. McAdams et al (74, 75) reported that long-term (over 3 years) treatment with alternate-day prednisone preserved renal function and improved glomerular morphology in children with both types of MPGN. Interestingly, dense deposit material remained unchanged in type II MPGN despite clinical and general histological improvement, analogous to the situation in most posttransplant recurrences. More recently, Donadio et al (76) reported stabilization of renal function and improved histology in patients treated with platelet inhibitor drugs (dipyridamole and aspirin).

Without well-controlled propsective studies, it is impossible to evaluate the efficacy of any therapy in diseases such as MPGN, well known for its extremely variable clinical course. Unfortunately, such studies are not currently available.

CONCLUSION

MPGN is one of the major causes of recognizable progressive glomerular annihilation and renal failure in man. Compelling evidence that this clinical-pathological syndrome represents a number of different entities and pathogenetic mechanisms is derived from studies carried out in the last decade. The common morphologic abnormality in this syndrome includes glomerular capillary wall thickening associated with mesangial expansion and cellular proliferation. Clinically, patients have varying degrees of hematuria and proteinuria, usually associated with mild nephrotic syndrome and hypertension, as well as varying degrees of progressive renal insufficiency.

Several morphologic types of MPGN have been identified. Type I is characterized by subendothelial deposits that contain immunoproteins—probably immune complex material. Type II has as its distinctive feature dense intramembranous and mesangial deposits of unknown antigenic composition, but they are surrounded by C3. And type III, which is less frequent, is characterized by the presence of subepithelial and subendothelial deposits containing immunoproteins. Both complement pathways are activated in type I whereas the alternative pathway is predominantly involved in type II disease. Although type II MPGN is included within the spectrum of MPGN, the presence of the intramembranous dense deposit, which regularly recurs in the transplanted kidney, characterizes this entity as being uniquely different from other forms of MPGN.

Although the pathogenesis of type I disease is unknown, the morphologic similarity to other forms of immune complex glomerulonephritis associated with certain infections suggests an infectious etiology. Currently, no specific therapy has been shown to alter definitively the natural history of this syndrome.

Literature Cited

1. West, C. D., McAdams, A. J., McConville, J. M., Davis, N. C., Holland, N. H. 1965. Hypocomplementemic and normocomplementemic persistent (chronic) glomerulonephritis; clinical and pathologic characteristics. *J. Pediatr.* 67:1089–1112
2. Gotoff, S. P., Fellers, F. X., Vawter, G. F., Janeway, C. A., Rosen, F. S. 1965. The beta 1 C globulin in childhood nephrotic syndrome. Laboratory diagnosis of progressive glomerulonephritis. *N. Engl. J. Med.* 273:524–29
3. Burkholder, P. M., Marchand, A., Krueger, R. P. 1970. Mixed membranous and proliferative glomerulonephritis. A correlative light, immunofluorescence, and electron microscopic study. *Lab. Invest.* 23:459–79
4. Burkholder, P. M., Hyman, L. R., Krueger, R. P. 1973. Characterization of mixed membranous and proliferative glomerulonephritis. Recognition of three varieties. In *Glomerulonephritis: Morphology, Natural History and Treatment,* ed. P. Kincaid-Smith, T. H. Mathew, E. L. Becker, Part 1:557–89. New York: Wiley. 653 pp.
5. Michael, A. F., Herdman, R. C., Fish, A. J., Pickering, R. J., Vernier, R. L. 1969. Chronic membranoproliferative glomerulonephritis with hypocomplementemia. *Transplant. Proc.* 1: 925–32
6. Jones, D. B. 1977. Membranoproliferative glomerulonephritis. One or many diseases? *Arch. Pathol. Lab. Med.* 101:457–61
7. Habib, R., Kleinknecht, C., Gubler, M. C., Levy, M. 1973. Idiopathic membranoproliferative glomerulonephritis in children. Report of 105 cases. *Clin. Nephrol.* 1:194–214
8. Droz, D., Zanetti, M., Noel, L.-H., Leibowitch, J. 1977. Dense deposit disease. *Nephron* 19:1–11
9. Cameron, J. S., Glasgow, E. F., Ogg, C. S., White, R. H. R. 1970. Membranoproliferative glomerulonephritis and persistent hypocomplementemia. *Br. Med. J.* 4:7–14
10. Habib, R., Gubler, M.-C., Loirat, C., Maiz, H. B., Levy, M. 1975. Dense deposit disease: a variant of membranoproliferative glomerulonephritis. *Kidney Int.* 7:204–15

11. Kim, Y., Vernier, R. L., Fish, A. J., Michael, A. F. 1979. Immunofluorescence studies of dense deposit disease. The presence of railroad tracks and mesangial rings. *Lab. Invest.* 40:474–80
12. Donadio, J. V. Jr., Slack, T. K., Holley, K. E., Ilstrup, D. M. 1979. Idiopathic membranoproliferative (mesangiocapillary) glomerulonephritis. A clinicopathologic study. *Mayo Clin. Proc.* 54:141–50
13. Davis, A. E., Schneeberger, E. E., Grupe, W. E., McCluskey, R. T. 1978. Membranoproliferative glomerulonephritis (MPGN Type I) and dense deposit disease (DDD) in children. *Clin. Nephrol.* 9:184–93
14. Michael, A. F., Westberg, N. G., Fish, A. J., Vernier, R. L. 1971. Studies on chronic membranoproliferative glomerulonephritis with hypocomplementemia. *J. Exp. Med.* 134:208s–27s
15. West, C. D. 1973. Membranoproliferative hypocomplementemic glomerulonephritis. *Nephron* 11:134–46
16. Antoine, B., Faye, C. 1972. The clinical course associated with dense deposits in the kidney basement membranes. *Kidney Int.* 1:420–27
17. Cameron, J. S., Ogg, C. S., Turner, D. R., Weller, R. O., White, H. R., Glasgow, E. F., Peters, D. K., Martin, A. 1973. Mesangiocapillary glomerulonephritis and persistent hypocomplementemia. See Ref. 4, 1:541–56
18. Herdman, R. C., Pickering, R. J., Michael, A. F., Vernier, R. L., Fish, A. J., Gewurz, H., Good, R. A. 1970. Chronic glomerulonephritis associated with low serum complement activity (chronic hypocomplementemic glomerulonephritis). *Medicine* 49:207–26
19. Northway, J. D., McAdams, A. J., Forristal, J., West, C. D. 1969. A silent phase of hypocomplementemic persistent nephritis detectable by reduced serum β1C-globulin levels. *J. Pediatr.* 74:28–38
20. Kincaid-Smith, P. 1973. The natural history and treatment of mesangiocapillary glomerulonephritis. See Ref. 4, 1:591–609
21. Cameron, J. S., Ogg, C. S., White, R. H. R., Glasgow, E. F. 1973. The clinical features and prognosis of patients with normocomplementemic mesangiocapillary glomerulonephritis. *Clin. Nephrol.* 1:8–13
22. Vargas, R., Thomson, K. J., Wilson, D., Cameron, J. S., Turner, D. R., Gill, D., Chantler, C., Ogg, C. S. 1976. Mesangiocapillary glomerulonephritis with dense deposits in the basement membranes of the kidney. *Clin. Nephrol.* 5:73–82
23. Lamb, V., Tisher, C. C., McCoy, R. C., Robinson, R. R. 1977. Membranoproliferative glomerulonephritis with dense intramembranous alterations. A clinicopathologic study. *Lab. Invest.* 36:607–17
24. Peters, D. K., Williams, D. G., Charlesworth, J. A., Boulton-Jones, J. M., Sissons, J. G. P., Evans, D. J. 1973. Mesangiocapillary nephritis, partial lipodystrophy, and hypocomplementemia. *Lancet* 2:535–38
25 Habib, R., Levy, M., Gubler, M.-C., Broyer, M., Habib, M. 1977. Lipodystrophie partielle, hypocomplementemie et glomerulonephrite. *Arch. Fr. Pédiatr.* 34:197–212
26. Sissons, J. G. P., West, R. J., Fallows, J., Williams, D. G., Boucher, B. J., Amos, N., Peters, D. K. 1976. The complement abnormalities of lipodystrophy. *N. Engl. J. Med.* 294:461–65
27. Bennett, W. M., Bardana, E. J., Wuepper, K., Houghton, D., Border, W. A., Götze, O., Schreiber, R. 1977. Partial lipodystrophy, C3 nephritic factor and clinically inapparent mesangiocapillary glomerulonephritis. *Am. J. Med.* 62: 757–60
28. Müller-Eberhard, H. J. 1975. Complement. *Ann. Rev. Biochem.* 44:697–724
29. Ooi, Y. M., Vallota, E. H., West, C. D. 1976. Classical complement pathway activation in membranoproliferative glomerulonephritis. *Kidney Int.* 9: 46–53
30. Habib, R., Loirat, C., Gubler, M.-C., Levy, M. 1974. Morphology and serum complement levels in membranoproliferative glomerulonephritis. *Adv. Nephrol.* 4:109–36
31. Spitzer, R. E., Vallota, E. H., Forristal, J., Sudora, E., Stitzel, A., Davis, N. C., West, C. D. 1969. Serum C'3 lytic system in patients with glomerulonephritis. *Science* 164:436–37
32. Vallota, E. H., Götze, O., Spiegelberg, H. L., Forristal, J., West, C. D., Müller-Eberhard, H. J. 1974. A serum factor in chronic hypocomplementemic nephritis distinct from immunoglobulins and activating the alternate pathway of complement. *J. Exp. Med.* 139:1249–61
33. Daha, M. R., Fearon, D. T., Austen, K. F. 1976. C3 nephritic factor (C3NeF): stabilization of fluid phase and cell-bound alternative pathway convertase. *J. Immunol.* 116:1–7

34. Davis, A. E. III, Ziegler, J. B., Gelfand, E. W., Rosen, F. S., Alper, C. A. 1977. Heterogeneity of nephritic factor and its identification as an immunoglobulin. *Proc. Natl. Acad. Sci. USA* 74:3980–83
35. Daha, M. R., Austen, K. F., Fearon, D. T. 1978. Heterogeneity, polypeptide chain composition and antigenic reactivity of C3 nephritic factor. *J. Immunol.* 120:1389–93
36. Scott, D. M., Amos, N., Sissons, J. G. P., Lachmann, P. J., Peters, D. K. 1978. The immunoglobulin nature of nephritic factor (NeF). *Clin. Exp. Immunol.* 32:12–24
37. Austen, K. F. 1978. Homeostasis of effector systems which can also be recruited for immunologic reactions. *J. Immunol.* 121:793–805
38. Williams, D. G., Peters, D. K., Fallows, J., Petrie, A., Kourilsky, O., Morel-Maroger, L., Cameron, J. S. 1974. Studies of serum complement in the hypocomplementemic nephritides. *Clin. Exp. Immunol.* 18:391–405
39. Levy, M., Gubler, M.-C., Sich, M., Beziau, A., Habib, R. 1978. Immunopathology of membranoproliferative glomerulonephritis with subendothelial deposits (type I MPGN). *Clin. Immunol. Immunopathol.* 10:477–92
40. Vallota, E. H., Forristal, J., Spitzer, R. E., Davis, N. C., West, C. D. 1971. Continuing C3 breakdown after bilateral nephrectomy in patients with membranoproliferative glomerulonephritis. *J. Clin. Invest.* 50:552–58
41. Peters, D. K., Williams, D. G. 1974. Complement and mesangiocapillary glomerulonephritis role of complement deficiency in the pathogenesis of nephritis. *Nephron* 13:189–97
42. Kim, Y., Shvil, Y., Michael, A. F. 1978. Hypocomplementemia in a newborn infant caused by placental transfer of C3 nephritic factor. *J. Pediatr.* 98:88–90
43. Ruddy, S., Carpenter, C. B., Chin, K. W., Knostman, J. N., Soter, N. A., Götze, O., Müller-Eberhard, H. J., Austen, K. F. 1975. Human complement metabolism: an analysis of 144 studies. *Medicine* 54:165–78
44. Peters, D. K., Martin, A., Weinstein, A., Cameron, J. S., Barratt, T. M., Ogg, C. S., Lachmann, P. J. 1972. Complement studies in membranoproliferative glomerulonephritis. *Clin. Exp. Immunol.* 11:311–20
45. Alper, C. A., Rosen, F. S., Watson, L. 1967. Studies of the in vivo behavior of human C'3 in normal subjects and patients. *J. Clin. Invest.* 46:2021–34
46. Colten, H. R., Levey, R. H., Rosen, F. S., Alper, C. A. 1973. Decreased synthesis of C3 in membranoproliferative glomerulonephritis. *J. Clin. Invest.* 52:20a (Abstr.)
47. Williams, D. G., Garcia-Fuentes, M., Losito, A. 1978. Cryoglobulinemia in mesangiocapillary glomerulonephritis. *Abstr. for 7th Int. Congr. Nephrol.,* Montreal, Canada, June 18–23, page G-11
48. Ooi, Y. M., Vallota, E. H., West, C. D. 1977. Serum immune complexes in membranoproliferative and other glomerulonephritides. *Kidney Int.* 11: 275–83
49. West, C. D., Winter, S., Forristal, J., McConville, J. M., Davis, N. C. 1967. Evidence for in vivo breakdown of β 1C-globulin in hypocomplementemic glomerulonephritis. *J. Clin. Invest.* 46:539–48
50. Kim, Y., Miller, K., Michael, A. F. 1977. Breakdown products of C3 and Factor B in hemolytic-uremic syndrome. *J. Lab. Clin. Med.* 89:845–50
51. Perrin, L. H., Lambert, P. H., Miescher, P. A. 1975. Complement breakdown products in plasma from patients with systemic lupus erythematosus and patients with membranoproliferative or other glomerulonephritis. *J. Clin. Invest.* 56:165–76
52. Charlesworth, J. A., Williams, D. G., Sherington, E., Lachmann, P. J., Peters, D. K. 1974. Metabolic studies of the third component of complement and the glycine-rich beta glycoprotein in patients with hypocomplementemia. *J. Clin. Invest.* 53:1578–87
53. West, C. D. 1976. Pathogenesis and approaches to therapy of membranoproliferative glomerulonephritis. *Kidney Int.* 9:1–7
54. West, C. D., McAdams, A. J. 1970. Serum β1C globulin levels in persistent glomerulonephritis with low serum complement: Variability unrelated to clinical course. *Nephron* 7:193–202
55. Vallota, E. H., Forristal, J., Davis, N. C., West, C. D. 1972. The C3 nephritic factor and membranoproliferative nephritis: correlation of serum levels of the nephritic factor with C3 levels, with therapy, and with progression of the disease. *J. Pediatr.* 80:947–59
56. Schürch, W., Leski, M., Hinglais, N. 1972. Evolution of recurrent lobular glomerulonephritis in a human kidney allotransplant. Combined light-, immunofluorescence-, and electron micro-

scopic studies of serial biopsies. *Virchows Arch. A* 355:66–84

57. Zimmerman, S. W., Hyman, L. R., Uehling, D. T., Burkholder, P. M. 1974. Recurrent membranoproliferative glomerulonephritis with glomerular properdin deposition in allografts. *Ann. Intern. Med.* 80:169–75
58. Berthoux, F. C., Ducret, F., Colon, S., Blanc-Brunat, N., Zech, P. Y., Traeger, J. 1975. Renal transplantation in mesangioproliferative glomerulonephritis (MPGN): relationship between the high frequency of recurrent glomerulonephritis and hypocomplementemia. *Kidney Int.* 7:S-323–S-327
59. McLean, R. H., Geiger, H., Burke, B., Simmons, R., Najarian, J., Vernier, R. L., Michael, A. F. 1976. Recurrence of membranoproliferative glomerulonephritis following kidney transplantation. Serum complement component studies. *Am. J. Med.* 60:60–72
60. Mathew, T. H., Mathews, D. C., Hobbs, J. B., Kincaid-Smith, P. 1975. Glomerular lesions after renal transplantation. *Am. J. Med.* 59:177–90
61. Hamburger, J., Crosnier, J., Noel, L.-H. 1978 Recurrent glomerulonephritis after renal transplantation. *Ann. Rev. Med.* 29:67–72
62. Porter, K. A., Dossetor, J. B., Marchioro, T. L., Peart, W. S., Rendall, J. M., Starzl, T. E., Terasaki, P. I. 1967. Human renal transplants. I. Glomerular changes. *Lab. Invest.* 16:153–81
63. Porter, K. A., Andres, G. A., Calder, M. W., Dossetor, J. B., Hsu, K. C., Rendall, J. M., Seegal, B. C., Starzl, T. E. 1968. Human renal transplants. II. Immunofluorescent and immunoferritin studies. *Lab. Invest.* 18:159–71
64. Mathew, T. H., Kincaid-Smith, P., Vikraman, P. 1977. Risks of vesicoureteric reflux in the transplanted kidney. *N. Engl. J. Med.* 297:414–18
65. Cameron, J. S., Turner, D. R. 1977. Recurrent glomerulonephritis in allografted kidneys. *Clin. Nephrol.* 7:47–54
66. Turner, D. R., Cameron, J. S., Bewick, M., Sharpstone, P., Melcher, D., Ogg, C. S., Evans, D. J., Trafford, A. J. P., Leibowitz, S. 1976. Transplantation in mesangiocapillary glomerulonephritis with intramembranous dense deposits: recurrence of disease. *Kidney Int.* 9:439–48
67. Galle, P., Hinglais, N., Crosnier, J. 1971. Recurrence of an original glomerular lesion in three renal allografts. *Transplant. Proc.* 3:368–70
68. Droz, D., Nabarra, B., Noel, L.-H., Leibowitch, J., Crosnier, J. 1979. Recurrence of dense deposits in transplanted kidneys: I. Sequential survey of the lesions. *Kidney Int.* 15:386–95
69. Leibowitch, J., Halbwachs, L., Wattel, S., Gaillard, M.-H., Droz, D. 1979. Recurrence of dense deposits in transplanted kidney: II. Serum complement and nephritic factor profiles. *Kidney Int.* 15:396–403
70. Kuriyama, T. 1973. Chronic glomerulonephritis induced by prolonged immunization in the rabbit. *Lab. Invest.* 28:224–35
71. Galle, P., Mahieu, P. 1975. Electron dense alteration of kidney basement membranes. A renal lesion specific of a systemic disease. *Am. J. Med.* 58: 749–64
72. Churg, J., Duffy, J. L., Bernstein, J. 1979. Identification of dense deposit disease. A report for the international study of kidney diseases in children. *Arch. Pathol. Lab. Med.* 103:67–72
73. Kincaid-Smith, P. 1972. The treatment of chronic mesangiocapillary (membranoproliferative) glomerulonephritis with impaired renal function. *Med. J. Aust.* 2:587–592
74. McAdams, A. J., McEnery, P. T., West, C. D. 1975. Mesangiocapillary glomerulonephritis: changes in glomerular morphology with long-term alternate day prednisone therapy. *J. Pediatr.* 86:23–31
75. McEnery, P. T. , West, C. D., McAdams, A. J. 1977. Membranoproliferative glomerulonephritis (MPGN): clinical follow-up of 36 children 1957 through 1977. *Abstr. for Am. Soc. Nephrol.*, Washington, DC, Nov. 20–22. page 18A
76. Donadio, J. V., Anderson, C. F., Mitchell, J. C., Fuster, V. D., Holley, K. E., Ilstrup, D. M. 1978. Controlled trial of platelet inhibitor drugs in membranoproliferative glomerulonephritis (MPGN). *Kidney Int.* 14:649 (Abstr.)

Ann. Rev. Med. 1980. 31:289–301
Copyright © 1980 by Annual Reviews Inc. All rights reserved

PATHOLOGICAL CHANGES IN AORTOCORONARY BYPASS GRAFTS

❖7361

Grover M. Hutchins, M.D.

The Department of Pathology, The Johns Hopkins Medical Institutions,
Baltimore, Maryland 21205

INTRODUCTION

The development of coronary artery bypass techniques (1–4) promised a direct solution to the problem of ischemic heart disease. It soon became apparent that the procedure provided relief or improvement of angina pectoris in 80–90% of patients (5–7). However, even now there is no clear evidence that patient survival (8, 9), ventricular function (10, 11), or control of arrhythmias (12, 13) is improved by the surgical procedure over results obtained by medical management. Further, there is even uncertainty that long-term relief of angina is sustained in operated patients. The means by which aortocoronary grafting provides relief of angina is puzzling because there is no evidence for similar frequencies of improvement in other aspects of ischemic heart disease. Several mechanisms have been suggested to account for the alleviation of pain, including increased blood flow, infarction of ischemic myocardium, and injury to cardiac nerves (14–17).

Studying the detailed morphological information from autopsied patients with coronary artery bypass grafts may improve our understanding of the procedure. Despite the many operations performed and the wealth of clinical follow-up data, there are relatively few reports of pathological changes in these patients, especially in long-term survivors. The complexity of the underlying coronary artery and myocardial disease and the difficulty of interpreting changes introduced by surgery have perhaps contributed to the paucity of pathological studies. Further, because the procedure carries a low operative mortality, the majority of patients do not die in the institution where the surgery was performed. This chapter discusses in relation to other studies the observations made on a series of patients with coronary artery bypass grafts autopsied at The Johns Hopkins Hospital. These patients' hearts were studied after postmortem arteriography.

0066-4219/80/0401-0289$01.00

AUTOPSY STUDY OF HEART

Proper autopsy studies of patients with ischemic heart disease are greatly facilitated by postmortem coronary arteriography. It is extremely difficult to adequately correlate coronary artery lesions and myocardial injuries without comparing radiographs with coronary arteries transected in situ and the topography of myocardial lesions with the coronary obstructions. Detailed histologic study may also be required to demonstrate the nature of the lesions in coronary arteries producing infarcts (18, 19). The interpretive problems are compounded if the patient has had cardiac surgery.

Postmortem Arteriography and Morphologic Examination

We have found the following technique (20) to be very useful in the autopsy examination of the hearts of patients with bypass grafts. Coronary arteries and any bypass grafts are injected with a barium-pigment-gelatin mass (21) at 120–150 mm Hg pressure. The heart is then distended with formalin at 20–40 cm H_2O pressure in the chambers. After overnight fixation stereoscopic radiographs (Figure 1) are made of the intact heart in at least two planes and of the transverse sections into which the heart is sliced.

Each saphenous vein to coronary artery anastomosis is removed as a block and radiographed. The block is decalcified, processed routinely, and serially sectioned at 8-μm intervals with every 15th section retained for histologic study. Sections of myocardium are taken for histology from the distribution of each graft, from any myocardial lesions noted, and from several random sites in the ventricular walls. The coronary arteries are transected at 2–3-mm intervals and portions of them, and of the grafts, are taken for histology as needed. Slides are routinely stained with hematoxylin and eosin and Verhoeff–van Gieson elastic.

The serial sections of each anastomosis are examined for the points of maximum obstruction related to the proximal and distal ends of the arteriotomy. The luminal area at these points is measured by planimetry and compared to the proximal or distal unoperated artery. Obstructions of either the proximal or distal arteriotomy site are considered significant if 75% or more of the area of the artery lumen is compromised.

Contraction Band Versus Coagulation Necrosis

An important histologic distinction to be made on myocardial injury is between contraction band necrosis and coagulation necrosis (18). Since the pathogenesis of the two lesions is different, their separation is helpful in interpretation of the case. Contraction band necrosis, also known as myofibrillar degeneration and hemorrhagic infarction, is characterized histologically by dense irregular eosinophilic clumping of the sarcoplasm with

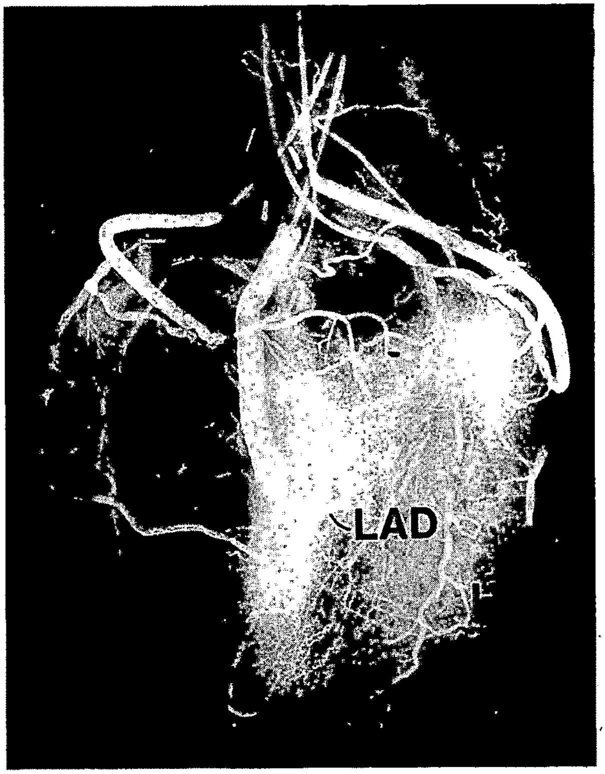


Figure 1 Postmortem coronary arteriogram of the heart of a 56-year-old man who underwent triple aortocoronary artery saphenous vein bypass grafting. Cannulae are still in place in the grafts and the left main coronary artery. Without angiography it would be difficult to appreciate the relationship of the graft to the artery and any occlusive lesions. In the circumflex graft the coronary artery at the distal end of the anastomosis is occluded while the proximal end is widely patent. The left anterior descending anastomosis has significant obstruction at both ends. The right anastomosis has been placed across a branch-point and one branch is occluded.

intervening cleared areas. Grossly, areas of contraction band necrosis are congested or hemorrhagic as a result of vascular injury accompanying the muscle cell necrosis. Contraction band necrosis arises from reperfusion into myocardium that has been subjected to a period of no perfusion. In contrast, coagulation necrosis preserves the general cytologic features of the tissue, is usually pale grossly, and arises from fixed obstruction to blood flow.

Patient Analysis

Upon completion of the pathological studies the entire case is reviewed including clinical information, autopsy findings, and the results of radiographic, gross, and histologic study of the coronary arteries, bypass grafts, and myocardium. The severity of overall and regional coronary artery disease is determined from the radiographs, gross examination, and histologic study. The type, extent, and age of myocardial injury is determined from gross and histologic examination.

So far we have studied 106 patients with coronary artery bypass grafts at autopsy with the techniques described above. These patients ranged in age from 35 to 75 years and 75 were male; 28 died at operation, 60 within one month and 18 between 1 and 70 months. There were 190 grafts placed in these patients including five internal mammary artery to left anterior descending coronary artery anastomoses. In the pathology of coronary artery bypass graft surgery it is useful to consider separately the graft itself, its connections, and the state of the myocardium in the distribution of grafted coronary arteries.

THE VEIN GRAFT

Clinical follow-up studies show that approximately 90% of grafts are patent in the operative period, but there is additional graft attrition over the course of time so that approximately two thirds remain patent after several years (10, 22, 23). In our 106 patients studied at autopsy 68 of 69 (99%) grafts were patent in patients dying at operation, 14 of 125 (91%) in those dying up to 30 days post operation, and 20 of 33 (61%) in those dying between 1 and 70 months. Histologic studies of grafts have identified two principal reactions, thrombosis and a concentric laminar deposition of intimal fibrous tissue (24–34). The frequency with which this intimal thickening, which is histologically the same as atherosclerosis, progresses to cause complete graft closure is uncertain.

Graft Thrombosis

Pathologic study of grafts that are completely closed after surgery shows typical thrombosis with organization and recanalization of the thrombus appropriate to its age. In many instances thrombosis of the graft may be ascribed to occlusion of the vein to coronary artery anastomosis; and in this circumstance, the thrombus is simply propagated from the arteriotomy site. A few grafts show complete thrombosis with a widely patent anastomotic site (Figure 2). The reason for this phenomenon is not entirely clear but may relate to a tendency for such grafted arteries to have less severe proximal obstructions or a reduced drainage bed as the result of infarction (35). Either hemodynamic situation would reduce the gradient across the graft and increase its tendency to thrombose.

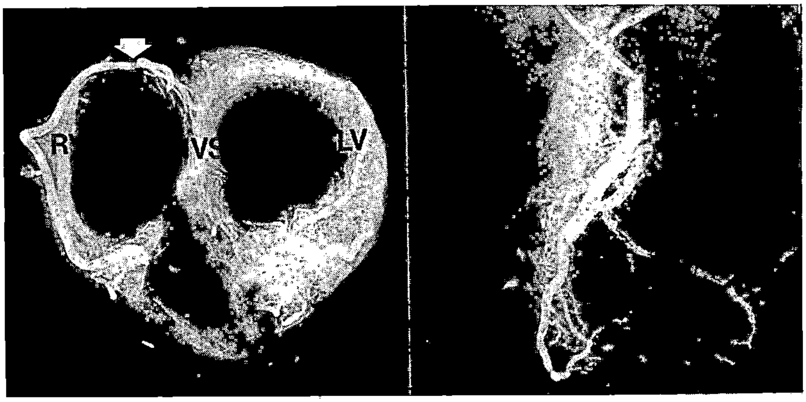


Figure 2 Comparison of occluded and patent aortocoronary saphenous vein bypass grafts. On the left is the postmortem radiograph of the base of the heart of a 70-year-old male who received a right coronary graft 4½ years prior to death. The graft-coronary artery anastomosis (arrow) is into the distal right coronary artery. The graft is completely occluded but the right coronary artery and anastomosis site show no significant obstruction. On the right a 3½-year-old graft into the left anterior descending coronary artery of a 68-year-old man is widely patent and smooth and uniform throughout. The artery proximal to the anastomosis site is totally occluded and the other coronary arteries have severe obstructions. These two cases illustrate the general tendency for grafts to close when flow through them is impeded, as by being placed into an artery with normal flow, and to remain patent when flow is favored, as with occlusion of the proximal artery and a large runoff through collateral vessels.

Other more specific causes of graft occlusion have been demonstrated or suggested. Infection of the graft with subsequent thrombosis has been but rarely encountered considering the relative frequency of postoperative mediastinitis (36, 37). Aneurysms of saphenous vein bypass grafts are rare. Diffuse dilatation in association with severe "atherosclerosis" of the grafts in two patients (38), and mycotic aneurysm of a vein graft (39), have been noted. Aneurysmal dilatation of the proximal portion of the graft with thrombotic occlusion of the vein beyond (40) may simply represent the result of a local injurious effect of organizing thrombus on the integrity of the vessel wall (41).

Abnormal tension on vein grafts has been described in 6 of 41 grafts examined in one series (42). This assessment was made in specimens not studied after postmortem arteriography and fixation in distention. We have not encountered this problem in our series of hearts where the vein grafts when injected are found to distend readily even though appearing flattened on initial examination. Indeed, the opposite problem, an excess of graft, appears to be a more frequent source of trouble. Jump grafts have a tendency to kink or roll thereby producing focal obstruction at the point where the graft is fixed by the side-to-side anastomosis (43). Fortunately, this complication is rare.

Concentric Fibrous Plaque

With patent grafts past the operative period a ubiquitous change is the thickening of the intima of the vein. This material consists of fibrin and mesenchymal cells and connective tissue fibers. In some grafts implanted for long periods calcium deposits, lipid-containing macrophages, and cholesterol crystals may be observed, i.e. changes typical of atherosclerosis. The origin of this concentric fibrous plaque is the conversion into connective tissue of thin laminae of thrombus deposited on the vein intima (31). The layer of fibrin is found in most grafts observed in the early hours after implantation. The material tends to be accentuated over the venous valves and at the lines of anastomosis of the graft to the aorta and coronary artery.

With passage of time the amount of connective tissue elements increases. However, even in grafts in place for long periods it is frequently possible to demonstrate fibrin deposited on the luminal surface of the concentric fibrous plaque. This observation suggests that progression of the process could lead to reduction in lumen size and possibly to late obliteration of the graft. It has been suggested that the intimal thickening represents "arterialization," a reaction to the increased pressure to which the graft is exposed (25). However, in other situations in which mural tension is increased as a result of hypertension or dilatation, the stereotypic intimal or endocardial response is fibroelastosis (44–46). The morphology of fibroelastosis, whether in the heart or systemic and pulmonary vessels, resembles the media of elastic arteries, having laminae of elastic tissue separated by mesenchymal cells. This is not the appearance of the intimal thickening seen in vein grafts, which usually has only a scanty complement of randomly arranged elastic fibers.

The evolution of the intimal change in the vein graft is similar to the sequence of histologic changes found in atherosclerosis. Duguid pointed out that thin intimal fibrin deposits occurred frequently on atherosclerotic plaques and that plaques grow by incorporating this material (47). The resemblance of the intimal change in vein grafts to atherosclerosis has been noted in several studies (29, 31, 32, 34). In occasional late grafts the intimal process may cause focal narrowings with a histologic resemblance to an advanced atherosclerotic plaque. It is likely that such focal accentuations of the intimal change occur where there are variations in regional hemodynamics such as points of tethering of the graft, where it changes course abruptly or is impinged on by adjacent structures (Figure 3).

One major factor in determining the thickness to which fibrous plaque develops may be the downstream runoff, that is, the flow volume passing

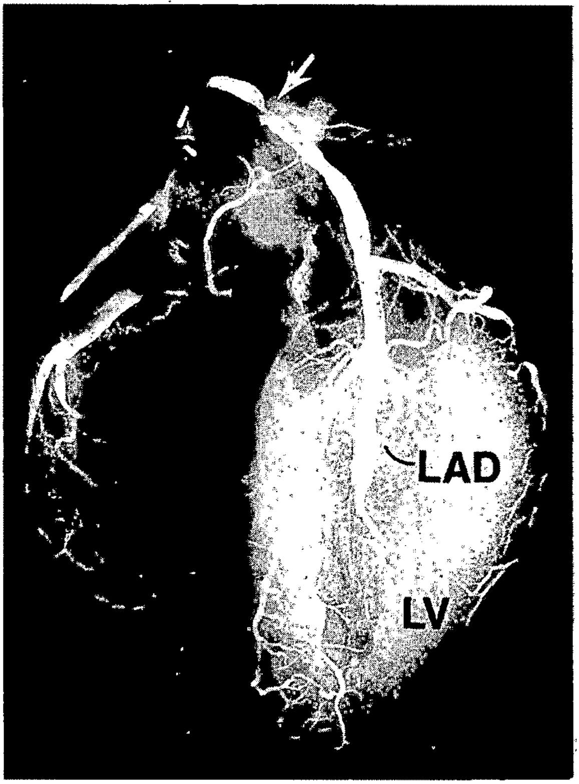


Figure 3 Postmortem coronary arteriogram from a 49-year-old man who received saphenous vein grafts to the distal right and mid left anterior descending coronary arteries five months prior to death. The right graft is occluded although part of its mid portion is filled by channels of recanalization in the organized thrombus. The left graft has a small region of marked narrowing (*arrow*) in its proximal part. Histologically this proved to be a typical atheroma. The anastomosis shows a branch-point problem with a patent diagonal and occluded left anterior descending coronary artery continuation. Interpretation of such a complex vascular bed is greatly facilitated by postmortem angiograms.

through the graft. Widely patent vein grafts with thin layers of fibrous plaque tend to occur where the anastomosis is patent and the proximal coronary artery occluded. This phenomenon may be clearly seen in occasional patients in whom the degree of occlusion of the coronary artery anastomosis varies (48).

Thus, the vein graft itself tends to either thrombose early in the operative period or to remain patent but develop a concentric layer of intimal fibrous tissue derived from the deposition of thin laminae of thrombus. In some grafts this intimal "atherosclerosis" becomes focally accentuated and may lead to graft occlusion. Both processes, the early occlusive thrombosis and the later progressive nonocclusive layering by thrombus appear to be enhanced by any process that reduces graft flow.

THE AORTA-VEIN GRAFT ANASTOMOSIS

Abnormalities of the proximal anastomosis appear to be infrequent. We have noted one patient with severe focal narrowing of a vein graft by intimal plaque near the aortic anastomosis. Two grafts in separate patients have shown aneurysmal dilatation of the proximal centimeter of the graft in association with complete thrombosis of the remainder of the graft, changes similar to a reported case (40). One other patient in our series (39) died as a result of dehiscence of a proximal graft in a setting of severe mediastinitis and mycotic aneurysm of the vein graft. Thus, the proximal or aorta to vein graft anastomosis appears to be relatively immune to any specific local pathologic changes.

THE VEIN GRAFT–CORONARY ARTERY ANASTOMOSIS

Pointed morphologic study of the distal anastomosis employing radiography and serial histologic sections has shown that compression, thrombosis, and dissection are the major mechanisms accounting for occlusion or significant obstruction of the anastomosis (45). Compression, which accounts for 40% of such abnormalities, occurs predominantly at the proximal and distal ends of the arteriotomy. The lumen may be reduced at these points through utilization of more of the arterial wall circumference for the anastomosis than is compensated for by the inserted vein graft wall. The problem is encountered more frequently in small arteries, since the suture bite necessarily uses a greater circumference of artery wall, and in anastomoses placed over atherosclerotic plaques.

As noted above, formation of a layer of thrombus on the intimal aspect of the graft and anastomosis is a normal and expected process that smoothly lines the surgically created conduit. In about 40% of those grafts beyond the operative period, excessive thrombosis obstructs the grafted artery at the anastomosis. The least common mechanism, causing about 8% of artery obstructions at the anastomosis, is dissection. This occurs as a result of injury to the arterial media. In 12% of anastomoses the obstruction is secondary to both compression and thrombosis. Obstructions may occur independently at either end of the arteriotomy site or both ends may be involved.

The same types of occlusive lesions have been noted in the side-to-side anastomoses of jump grafts as in the more common end-to-side anastomoses (43). We have noted also that anastomoses placed so that the arteriotomy extends into a branch artery are frequently obstructed in either the grafted or the nongrafted branch artery. When the arteriotomy was placed so as to end just proximal to the flow divider of the branch-point, occlusion was not observed. Apparently the tendency for arteries with small caliber or atheromata at the anastomosis site (ostial lesions are the usual reason for extending the arteriotomy into the branch) to develop obstruction accounts for this observation.

Thus, the occlusions observed in the coronary arteries following bypass grafting (50) appear to be predominantly related to the technical aspects of creating the anastomosis.

THE MYOCARDIUM IN THE GRAFT DISTRIBUTION

Two explanations have been proposed to account for the relief of angina after bypass graft surgery: blood flow is increased to ischemic myocardium, or the pain-producing segment of myocardium is infarcted (16, 17). The unexpected observation that emerged from pathologic studies of myocardium in the distribution of grafted coronary arteries was that regional myocardial necrosis tended to occur when the graft was widely patent and that myocardium with obstructed revascularization tended to be spared.

Those vein to coronary artery anastomoses that were placed across branch points clearly demonstrated the relationship between patency and myocardial injury (51). Five of six grafts with different degrees of occlusion in the two branch vessels showed a greater degree of necrosis with the more patent branch. The injured myocardium showed contraction band necrosis, the form of injury that occurs when myocardium is reperfused following a period of no perfusion. This histologic form of myocardial injury must be distinguished from coagulation necrosis, which is produced by fixed obstructions to coronary blood flow. Clearly, observation of contraction band necrosis in the unobstructed branch vessel indicated the myocardium was sufficiently altered during the anoxic arrest of surgery to permit injury to develop when reperfusion was brisk. The surviving myocardium in the distribution of the obstructed branch obtained sufficient reflow to remain alive, probably via collaterals, but at such a rate as to permit the cells to recover from the changes occurring during the same period of nonperfusion. A similar phenomenon is seen in the stone heart syndrome where massive myocardial contraction band necrosis occurs in hearts with widely patent coronary arteries (52). Rapid unimpeded reperfusion following the period of myocardial arrest to accomplish valve replacement produces the contraction band necrosis in these cases.

Examination of the hearts of 58 patients dying within one month of receiving bypass grafts (53) demonstrated that 22 (38%) had regional transmural necrosis in the distribution of grafted coronary arteries. Of the regional necroses 18 were of the contraction band or reperfusion type and 15 of them occurred in the distribution of a patent graft to a coronary artery. There were four areas of coagulation, or no-reflow, necrosis and all of these were in the distribution of occluded graft systems. This study showed that while regional transmural necrosis was frequent it occurred predominantly in association with graft patency and that while graft obstruction was frequent it seldom led to myocardial injury. The frequency of necrosis observed in these patients was similar to that detected in some clinical studies (16). Other pathologic studies have also noted a high incidence of myocardial necrosis in patients dying after graft implantation (54, 55).

The reports of myocardial injury following cardiac operation under cardiopulmonary bypass described above were mainly in patients where the myocardium was rendered hypothermic and then given anoxic arrest. More recently the use of potassium chloride solution to arrest the heart prior to aortic cross-clamping appears to improve ventricular function and reduce the incidence of operatively related myocardial injury (56–58). In patients where cold potassium chloride solution arrest was employed at operation and who later came to autopsy, the extent of contraction band necrosis in the distribution of grafted arteries is significantly reduced.

Thus, the pathologic evidence suggests that the correlation between patency of a graft system and the state of the myocardium in its distribution has paradoxical features in many patients. With many widely patent grafts the myocardium has undergone contraction band, or reperfusion, necrosis. With most occluded grafts the myocardium has shown no morphologic effect, probably a tribute to the effective collateral flow in these patients. It is hoped that with recognition of these effects modified operative techniques will enhance myocardial preservation.

SUMMARY

Despite more than 10 years of experience with coronary artery bypass graft surgery, a number of basic pathophysiologic questions remain unanswered. Some, such as whether or not the procedure prolongs life, improves ventricular function, and reduces the incidence of arrhythmias, must be answered through controlled prospective clinical trials. The solution of other problems such as the pathophysiology of angina and the mechanism of its relief by coronary artery surgery, the complications of cardiac surgery, and the long-term fate of grafts, grafted coronary artery systems, and the myocardium require data derived from detailed morphologic studies of autopsied patients.

Literature Cited

1. Garrett, H. E., Dennis, E. W., DeBakey, M. E. 1973. Aortocoronary bypass with saphenous vein graft. Seven-year follow-up. *J. Am. Med. Assoc.* 223:792–94
2. Favaloro, R. G. 1968. Saphenous vein autograft replacement of severe segmental coronary artery occlusion. Operative technique. *Ann. Thorac. Surg.* 5:334–39
3. Kerth, W. J. 1969. Aortocoronary bypass grafts. *J. Thorac. Cardiovasc. Surg.* 57:487–92
4. Johnson, W. D., Flemma, R. J., Lepley, D. Jr., Ellison, E. H. 1969. Extended treatment of severe coronary artery disease: A total surgical approach. *Ann. Surg.* 170:460–70
5. Favaloro, R. G., Effler, D. B., Groves, L. K., Sheldon, W. C., Shirey, E. K., Sones, F. M. Jr. 1970. Severe segmental obstruction of the left main coronary artery and its divisions. Surgical treatment by the saphenous vein graft technique. *J. Thorac. Cardiovasc. Surg.* 60: 469–82
6. Reul, G. J.Jr., Cooley, D. A., Wukasch, D. C., Kyger, E. R. III, Sandiford, F. M., Hallman, G. L., Norman, J. C. 1975. Long-term survival following coronary artery bypass. Analysis of 4,522 consecutive patients. *Arch. Surg.* 110: 1419–24
7. Najmi, M., Ushiyama, K., Blanco, G., Adam, A., Segal, B. L. 1974. Results of aortocoronary artery saphenous vein bypass surgery for ischemic heart disease. *Am. J. Cardiol.* 33:42–48
8. Murphy, M. L., Hultgren, H. N., Detre, K., Thomsen, J., Takaro, T., participants of Veterans Administration Cooperative Study. 1977. Treatment of chronic stable angina. A preliminary report of survival data of the randomized Veterans Administration Cooperative Study. *N. Engl. J. Med.* 297:621–27
9. McIntosh, H. D., Garcia, J. A. 1978. The first decade of aortocoronary bypass grafting, 1967–1977. A review. *Circulation* 57:405–31
10. Lawrie, G. M., Morris, G. C. Jr., Howell, J. F., Ogura, J. W., Spencer, W. H. III, Cashion, W. R., Winters, W. L., Beazley, H. L., Chapman, D. W., Peterson, P. K., Lie, J. T. 1977. Results of coronary bypass more than 5 years after operation in 434 patients. *Am. J. Cardiol.* 40:665–72
11. Hammermeister, K. E., Kennedy, J. W., Hamilton, G. W., Stewart, D. K., Gould, K. L., Lipscomb, K., Murray, J. A. 1974. Aortocoronary saphenous-vein bypass. Failure of successful grafting to improve resting left ventricular function in chronic angina. *N. Engl. J. Med.* 290:186–92
12. Tilkian, A. G., Pfeifer, J. F., Barry, W. H., Lipton, M. J., Hultgren, H. N. 1976. The effect of coronary bypass surgery on exercise-induced ventricular arrhythmias. *Am. Heart J.* 92:707–14
13. Guinn, G. A., Mathur, V. S. 1976. Surgical versus medical treatment for stable angina pectoris: Prospective randomized study with 1- to 4-year follow-up. *Ann. Thorac. Surg.* 22:524–27
14. Soloff, L. A. 1973. Effects of coronary bypass procedures. *N. Engl. J. Med.* 288:1302–3
15. Ross, R. S. 1975. Ischemic heart disease: an overview. *Am. J. Cardiol.* 36:496–505
16. Achuff, S. C., Griffith, L. S. C., Conti, C. R., Humphries, J. O., Brawley, R. K., Gott, V. L., Ross, R. S. 1975. The "angina-producing" myocardial segment: An approach to the interpretation of results of coronary bypass surgery. *Am. J. Cardiol.* 36:723–33.
17. Mnayer, M., Chahine, R. A., Raizner, A. E. 1977. Mechanisms of angina relief in patients after coronary artery bypass surgery. *Br. Heart J.* 39:605–9
18. Ridolfi, R. L., Hutchins, G. M. 1977. The relationship between coronary artery lesions and myocardial infarcts. Ulceration of atherosclerotic plaques precipitating coronary thrombosis. *Am. Heart J.* 93:468–86
19. Prizel, K. R., Hutchins, G. M., Bulkley, B. H. 1978. Coronary artery embolism and myocardial infarction: A clinicopathologic study of 55 patients. *Ann. Intern. Med.* 88:155–61
20. Hutchins, G. M., Anaya, O. A. 1973. Measurements of cardiac size, chamber

volumes and valve orifices at autopsy. *Johns Hopkins Med. J.* 133:96–106

21. Hales, M. R., Carrington, C. B. 1971. A pigmented gelatin mass for vascular injection. *Yale J. Biol. Med.* 43:257–70
22. Rimm, A. A., Blumlein, S., Barboriak, J. J., Anderson, A. J., Walker, J. A., Johnson, W. D. 1976. The probability of closure in aortocoronary vein bypass grafts. *J. Am. Med. Assoc.* 236:2637–40
23. FitzGibbon, G. M., Burton, J. R., Leach, A. J. 1978. Coronary bypass graft fate. Angiographic grading of 1400 consecutive grafts early after operation and of 1132 after one year. *Circulation* 57:1070–74
24. Grondin, C. M., Meere, C., Castonguay, Y., Lepage, G., Grondin, P. 1971. Progressive and late obstruction of an aorto-coronary venous bypass graft. *Circulation* 43:698–702
25. Vlodaver, Z. Edwards, J. E. 1971. Pathologic changes in aortic-coronary arterial saphenous vein grafts. *Circulation* 44:719–28
26. Marti, M. C., Bouchardy, B., Cox, J. N. 1971. Aorto-coronary by-pass with autogenous saphenous vein grafts: Histopathological aspects. *Virchows Arch. A.* 352:255–66
27. Kern, W. H., Dermer, G. B., Lindesmith, G. G. 1972. The intimal proliferation in aortic-coronary saphenous vein grafts. Light and electron microscopic studies. *Am. Heart J.* 84:771–77
28. Unni, K. K., Kottke, B. A., Titus, J. L., Frye, R. L., Wallace, R. B., Brown, A. L. 1974. Pathologic changes in aortocoronary saphenous vein grafts. *Am. J. Cardiol.* 34:526–32
29. Barboriak, J. J., Pintar, K., Korns, M. E. 1974. Atherosclerosis in aortocoronary vein grafts. *Lancet* 2:621–24
30. Lawrie, G. M., Lie, J. T., Morris, G. C. Jr., Beazley, H. L. 1976. Vein graft patency and intimal proliferation after aortocoronary bypass: Early and long-term angiopathologic correlations. *Am. J. Cardiol.* 38:856–62
31. Bulkley, B. H., Hutchins, G. M. 1977. Accelerated "atherosclerosis." A morphologic study of 97 saphenous vein coronary artery bypass grafts. *Circulation* 55:163–69
32. Lie, J. T., Lawrie, G. M., Morris, G. C. Jr. 1977. Aortocoronary bypass saphenous vein graft atherosclerosis. Anatomic study of 99 vein grafts from normal and hyperlipoproteinemic patients up to 75 months post-operatively. *Am. J. Cardiol.* 40:906–14
33. Barboriak, J. J., Pintar, K., VanHorn, D. L., Bataylas, G. E., Korns, M. E. 1978. Pathologic findings in the aortocoronary vein grafts. A scanning electron microscope study. *Atherosclerosis* 29:69–80
34. Grondin, C. M., Campeau, L., Lesperance, J., Solymoss, B. C., Vouhe, P., Castonguay, Y. R., Meere, C., Bourassa, M. G. 1979. Atherosclerotic changes in coronary vein grafts six years after operation. Angiographic aspect in 110 patients. *J. Thorac. Cardiovasc. Surg.* 77:24–31
35. Kennedy, J. H., Wieting, D. W., Hwang, N. H. C., Anderson, M. S., Bayardo, R. J., Howell, J. F., DeBakey, M. E. 1974. Hydraulic and morphologic study of fibrous intimal hyperplasia in autogenous saphenous vein bypass grafts. *J. Thorac. Cardiovasc. Surg.* 67:805–13
36. Macmanus, Q., Okies, J. E. 1976. Mediastinal wound infection and aortocoronary graft patency. *Am. J. Surg.* 132:558–61
37. Bulkley, B. H., Hutchins, G. M. 1978. Acute postoperative graft phlebitis: A rare cause of saphenous vein-coronary artery bypass failure. *Am. Heart J.* 95:757–60
38. Pintar, K., Barboriak, J. J., Johnson, W. D., Co, E. D. 1978. Atherosclerotic aneurysm in aortocoronary vein graft. *Arch. Pathol. Lab. Med.* 102:287–88
39. Douglas, B. R., Bulkley, B. H., Hutchins, G. M. 1979. Infected saphenous vein coronary artery bypass graft with mycotic aneurysm and fatal dehiscence of the proximal anastomosis. *Chest* 75:76–77
40. Riahi, M., Vasu, C. M., Tomatis, L. A., Schlosser, R. J., Zimmerman, G. 1975. Aneurysm of saphenous vein bypass graft to coronary artery. *J. Thorac. Cardiovasc. Surg.* 70:358–59
41. Salyer, W. F., Salyer, D. C., Hutchins, G. M. 1974. Local arterial wall injury caused by thromboemboli. *Am. J. Pathol.* 75:131–40
42. Spray, T. L., Roberts, W. C. 1976. Tension on coronary bypass conduits. A neglected cause of real or potential obstruction of saphenous vein grafts. *J. Thorac. Cardiovasc. Surg.* 72:282–87
43. Hutchins, G. M., Bulkley, B. H. 1977. Mechanisms of occlusion of saphenous vein-coronary artery "jump" grafts. *J. Thorac. Cardiovasc. Surg.* 73:660–67
44. Hutchins, G. M., Ostrow, P. T. 1976. The pathogenesis of the two forms of

hypertensive pulmonary vascular disease. *Am. J. Med.* 61:361–66

45. Hutchins, G. M., Bannayan, G. A. 1971. Development of endocardial fibroelastosis following myocardial infarction. *Arch. Pathol.* 91:113–18
46. Hutchins, G. M., Vie, S. A. 1972. The progression of interstitial myocarditis to idiopathic endocardial fibroelastosis. *Am. J. Pathol.* 66:483–96
47. Duguid, J. B. 1946. Thrombosis as a factor in the pathogenesis of coronary artherosclerosis. *J. Pathol. Bacteriol.* 58:207–12
48. Bulkley, B. H., Humphries, J. O., Hutchins, G. M. 1977. Purulent pericarditis with asymmetric cardiac tamponade: A cause of death months after coronary artery bypass surgery. *Am. Heart J.* 93:776–83
49. Griffith, L. S. C., Bulkley, B. H., Hutchins, G. M., Brawley, R. K. 1977. Occlusive changes at the coronary artery-bypass graft anastomosis. *J. Thorac. Cardiovasc. Surg.* 73:668–79
50. Griffith, L. S. C., Achuff, S. C., Conti, C. R., Humphries, J. O., Brawley, R. K., Gott, V. L., Ross, R. S. 1973. Changes in intrinsic coronary circulation and segmental ventricular motion after saphenous-vein coronary bypass graft surgery. *N. Engl. J. Med.* 288: 590–95
51. Hutchins, G. M., Bulkley, B. H. 1977. Correlation of myocardial contraction band necrosis and vascular patency. *Lab. Invest.* 36:642–48
52. Hutchins, G. M., Silverman, K. J. 1979. Pathology of the stone heart syndrome: Massive myocardial contraction band necrosis and widely patent coronary arteries. *Am. J. Pathol.* 95:745–52
53. Bulkley, B. H., Hutchins, G. M. 1977. Myocardial consequences of coronary artery bypass graft surgery. The paradox of necrosis in areas of revascularization. *Circulation* 56:906–13
54. Vlodaver, Z., Edwards, J. E. 1973. Pathologic analysis in fatal cases following saphenous vein coronary arterial bypass. *Chest* 64:555–63
55. Heard, B. E. 1976. Pathology of hearts after aortocoronary saphenous vein bypass grafting for coronary artery disease, studied by postmortem coronary angiography. *Br. Heart J.* 38:839–59
56. Gay, W. A. Jr., Ebert, P. A. 1973. Functional, metabolic, and morphologic effects of potassium induced cardioplegia. *Surgery* 74:284
57. Schaff, H. V., Dombroff, R., Flaherty, J. T., Bulkley, B. H., Hutchins, G. M., Goldman, R. A., Gott, V. L. 1978. Effect of potassium cardioplegia on myocardial ischemia and post arrest ventricular function. *Circulation* 58: 240–49
58. Gharagozloo, F., Bulkley, B. H., Hutchins, G. M., Bixler, T. J. II, Schaff, H. V., Flaherty, J. T., Gardner, T. J. 1979. Potassium cardioplegia during normothermic cardiac arrest: A morphologic study of the effect of varying concentrations of potassium on myocardial anoxic injury. *J. Thorac. Cardiovasc. Surg.* 77:602–7

Ann. Rev. Med. 1980. 31:303–14
Copyright © 1980 by Annual Reviews Inc. All rights reserved

ERYTHROPOIESIS[1]

◆7362

C. Peschle, M.D.

Institute of Medical Pathology and ICHO, II Faculty of Medicine and Surgery,
University of Naples, Naples, Italy

In the 1960s and early 1970s, studies on ^{59}Fe- and ^{3}H-thymidine-labeled erythroid cells allowed insight into the physiology of erythropoietin and the kinetics of erythroblasts (1–3). However, little information was available on the progenitor cells of erythroblasts. In vitro assays for these precursors (4) have recently been developed, and we can now explore their kinetics and control mechanisms, thus partially unveiling the intimate phenomena underlying erythropoietic differentiation. This review focuses on these recent advances (cf 1–7).

CLONOGENIC ASSAYS OF HEMOPOIETIC PRECURSORS

In these semisolid culture assays a progenitor cell, triggered by a hemopoietic hormone, gives rise to a clonal colony, i.e. a cell progeny kept together in a network of agar, fibrin, or methylcellulose (4). The number of colonies depends upon both the number of progenitors and the hormonal activity. With these techniques we can identify the precursors of the various hemopoietic lineages (erythroid, myeloid, monocytic, eosinophilic, megakaryocytic, B, and T lymphocytic) (4). They require, however, elevated concentrations of fetal calf serum (FCS), which modulates colony growth via obscure mechanisms, thus confusing both methodology and interpreta-

---

[1] Abbreviations used in this chapter are as follows:
CFU-S: colony-forming unit, spleen.
BFU-E: burst-forming unit, erythroid.
CFU-E: colony-forming unit, erythroid.
CFU-GM: colony-forming unit, granulomonocytic.
BPA: burst-promoting activity; also called burst-enhancing factor (BEF).
Ep: erythropoietin.
CSF: colony-stimulating factor.

0066-4219/80/0401-0303$01.00

tion. Iscove (8) recently developed a clonogenic system almost totally devoid of FCS, which is effectively replaced by its active constituents. Further limitations in these assays derive from the impurity of the routine preparations of hemopoietic hormones (Ep and CSF), as well as the possibility of in vitro interactions between assayed precursors and other cell systems (see below).

THE ERYTHROID PRECURSORS

Pluripotent hemopoietic stem cells are capable of both self-renewal and differentiation to various precursors which are "committed" to erythroid, myelomonocytic, megakaryocytic, or lymphoid lineages (4–7). These progenitors in turn give rise to the respective differentiated cell series.

The murine pluripotent element is assayed on the basis of its capacity to form mixed spleen colonies (CFU-S) in sublethally radiated mice (9). An in vitro clonogenic assay for murine CFU-S is being developed by Metcalf & Johnson (10, 11).

Two erythroid progenitors have been identified in mammals, the erythroid burst- (BFU-E) and colony-forming (CFU-E) units, which give rise respectively to large "bursts"[2] or small colonies after Ep addition (12). The clonal origin of CFU-E aggregates has been demonstrated by G-6-PD isoenzyme patterns (13, see also the section on Clinical Aspects), that of bursts is indicated by linearity of log dose/response regression lines for both Ep and BPA (8, see below).

Since erythroid bursts are largely composed of CFU-E colonies, BFU-E clearly represent early erythroid progenitors, which differentiate into late precursors, i.e. CFU-E.

It is generally conceded that BFU-E are closely related to the CFU-S compartment. Thus, BFU-E and CFU-S are characterized by identical cell size and buoyant density (14), although the BFU-E shows a more elevated cycling activity than the slowly proliferating CFU-S (15–17). BFU-E–CFU-E differentiation is characterized by a further increase of proliferative rate (16, 17), as well as of pool amplification, cell size, and density (14, 16, 17). Finally, CFU-E feed into the erythroblastic pool (4–7). In this regard, the kinetics of CFU-E and pro-erythroblasts partially overlap, which suggests a close relationship between these elements (18).

Differentiation of erythroid precursors thus entails a gradual increase of their number, proliferative activity, cell size, and density. A further differential marker is the progressive enhancement of their Ep sensitivity (16, 17), possibly reflecting an increase in the number of Ep receptors on the membranes of differentiating progenitors (see below).

---

[2] The term burst refers to the rapid growth of these colonies after an initial period of quiescence or slow proliferation.

In both humans and mice, Eaves et al (17) identified a third erythroid precursor, the "mature" BFU-E, which is apparently intermediate between "primitive" BFU-E and CFU-E with respect to all the above parameters. It therefore seems evident that erythroid progenitors are characterized by a continuous spectrum of differentiation stages.

A circulating BFU-E has been identified in murine blood (19), as well as in the null lymphocyte fraction of human blood (20).

REGULATION OF ERYTHROPOIESIS

Ep Control of Erythropoiesis

The erythropoietic rate is largely or solely regulated by Ep (1–3), which has recently been purified (21).

Ep specifically induces differentiation of the late erythroid precursors to the recognizable erythroid compartment (22). At the molecular level this effect is apparently mediated via (*a*) interaction of Ep with a putative membrane protein receptor (cf 23); (*b*) activation of a cytoplasmic factor transferring the hormonal message to the genome (23); (*c*) transcription of specific RNA molecules (24); and (*d*) protein specialization (synthesis of Hb, spectrin, membrane antigens, etc).

Ep action on erythroid precursors is further discussed below. Ep induces an accelerated maturation of erythroblasts (25); in stress erythropoiesis, enhanced Ep activity causes skipping of terminal divisions (in the presence of a relatively fixed generation time) and therefore macrocytosis (25).

Kinetics of Erythroid Precursors: Role of Ep and Burst-Promoting Activity

The CFU-E compartment is amplified in the face of elevated Ep levels, but depleted by suppressed Ep activity in transfusion-induced polycythemia (16, 18). The size of this pool is thus directly correlated with Ep activity. The relative number of CFU-E in DNA synthesis is slightly but significantly diminished in transfused mice (16, 26). These observations indicate that the kinetics of the CFU-E pool are largely (but not exclusively, see below) controlled by Ep levels.

In transfused mice, sharp depletion of the CFU-E pool (16, 18) and suppression of erythropoiesis (22) are surprisingly associated with a relatively elevated number of CFU-E in DNA synthesis (16, 26). The destiny of these proliferating CFU-E is uncertain. In this regard, Ep is required in vitro for CFU-E survival (8); marked decrease of Ep activity after transfusion may cause massive loss of actively proliferating CFU-E.

Starting from day 2–3 after erythroid stress or suppression, BFU-E cycling and pool size are in the normal range; the kinetics of these precursors is apparently insensitive to Ep levels (16). However, in the first 12–24 hr after erythroid perturbations, both the proliferative rate and hence the number of BFU-E are correlated directly with fluctuations of Ep activity (26, 27). Afterwards, both parameters return to nearly normal values (26, 27), in agreement with previous studies (16). BFU-E kinetics is thus apparently modulated by Ep activity, but only early after erythroid perturbations. The early Ep sensitivity is followed by an "escape" phenomenon, whereby BFU-E cycling and pool size become insensitive to Ep and return to normal steady-state values. The homeostatic factor underlying this escape is possibly represented by BPA (27).

Cultures of radiated murine marrow (14, 28) and medium conditioned by lectin-stimulated spleen cells (8, 10, 11) contain BPA, a non-Ep factor(s) promoting in vitro formation of erythroid bursts. Detection of BPA is rendered difficult by its presence in FCS, which is a standard constituent of culture medium. However, this difficulty may be circumvented by either employing FCS low in BPA or reducing FCS content in the medium (8). In these modified culture conditions, BPA allows survival of BFU-E in the absence of Ep and permits burst formation in the presence of Ep (8). Furthermore, BPA enhances BFU-E cycling in vitro (28a). In humans, a BPA-like factor is present in peripheral leukocyte-conditioned medium (17, 29, 30) and a T-lymphocytes cell line (see below).

Preliminary evidence suggests that BPA levels fluctuate in serum (28), thus possibly modulating in vivo the kinetics of BFU-E and CFU-E pools to respectively a major and minor extent. As previously indicated, transfusion in mice induces an early decline of BFU-E cycling and number, then both parameters return to normal; this rebound is perhaps mediated by elevation of BPA (27). Furthermore, transfused mice subjected to radiation or hydroxyurea treatment and then injected with normal marrow cells show a wave of Ep-independent erythroid regeneration, which is characterized by enhanced BFU-E cycling (16) and an amplified CFU-E pool (18, 28a). These phenomena may also be modulated by a rise of BPA.

Biogenesis of Ep and BPA

The kidney is the major site of adult Ep production (31). The Ep molecule is synthesized (32) and secreted by the renal tissue (33), possibly via activation of a renal pro-Ep (34) [i.e. the renal erythropoietic factor, REF (35)]. The process may include an enzymatic activation step within the kidney (36). Ep production is apparently modulated by renal Po_2 (i.e. O_2 supply/demand ratio), via a putative renal O_2 sensor and the PGE-cAMP system (37). Studies with fluoresceinated anti-Ep serum suggested that

Ep-producing cells reside in the glomerulus (38). Further progress in this complex area clearly depends upon development of new technology, e.g. a sensitive specific radioimmunoassay for Ep and monovalent anti-Ep sera. Renal tissue cultures, although disappointing in the past, may prove of crucial value in the future.

In the mammalian fetus (and also the neonate rat), the liver is the major or sole source of Ep (39, 40). In the perinatal period renal Ep production begins (40). In the adult, liver Ep is masked by kidney Ep, i.e. the hepatic component becomes a minor aliquot of the total hormonal activity (41). Hepatic macrophages (Kupffer cells) are a source not only for CSF (42), but possibly also for Ep (43). Thus, their hyperplasia in anephric rats is associated with enhanced Ep response to hypoxia (41). Furthermore, subtotal liver excision is followed by elevation of hepatic Ep production (44, 45), which parallels macrophage regeneration (45). Neonatal Kupffer cells interact in vitro with fluoresceinated serum raised against partially purified Ep (46). In fetal liver erythropoiesis, these cells may therefore represent a local source of Ep. In adult marrow, macrophages are strategically located in the center of erythroblastic islands (47): although their function is unknown, they may play a key regulatory role, via BPA and PGE release, or even as a local source for Ep.

The mononuclear fraction of human blood, when deprived of T lymphocytes, sharply loses its capacity to form erythroid bursts (20). In this regard, T lymphocytes, and possibly macrophages release BPA, as indicated by presence of this factor in medium of cell lines of human T lymphocytes (48) and murine macrophages (WEHI-3 cells) (48a). Apparently, the T lymphocyte–macrophage complex may play a major role in BPA production, besides its function as a source of CSF (49, 50).

Competition Between Erythroid and Granulo-Monocytic Precursors: An Open Question

In vivo, the rates of granulopoiesis and erythropoiesis are inversely related (51, 52). In vitro, Ep and CSF compete for cloning of BFU-E and granulomonocytic precursors (CFU-GM) respectively, but only at high (53), not low (16), cell densities (respectively, $> 10^5$ or 10^4 cells per plate). Since the concentration of cells within the bone marrow cavity was estimated to be 1.8×10^9 per ml (54), high cell densities in vitro may be required to duplicate the in vivo conditions.

Hemopoietic Differentiation and Its Control Mechanisms

A spectrum of differentiating GM-progenitors has been demonstrated (55): these precursors are characterized by gradual increase of density and cell size, the early ones showing values respectively identical or similar to those of CFU-S (14, 55). Two types of CSF have been purified, a granulomonocytic CSF (55a) and a CSF that selectively stimulates monocytic growth (56).

Primitive BFU-E colonies are not composed exclusively of erythroid cells. A majority of them contain megakaryocytes (14), which suggests that primitive BFU-E are dipotent for erythroid and megakaryocytic differentiation. Upon addition of elevated amounts of BPA to murine cultures, a large proportion of bursts also contain granulocytes, i.e. constitute "mixed" colonies (28a). Similarly, human cultures stimulated sequentially by BPA and Ep show a significant number of "mixed" bursts (56a). BPA per se promotes GM-colony growth, whereas purified GM-CSF, in the presence of Ep, exerts a BPA-like action (11). A crucial question concerns, therefore, the possible identity of BPA and GM-CSF, and more generally the relationship between recently described factors acting on early hemopoietic precursors. In this regard, Wagemaker (28a) apparently separated from lectin-stimulated spleen-conditioned medium three distinct glycoprotein factors: (*a*) the first one enhances proliferation and differentiation of primitive BFU-E; (*b*) the second one promotes proliferation and differentiation of a primitive myeloid progenitor to progeny responding to GM-CSF; and (*c*) the third one induces proliferation of CFU-S.

A further matter of uncertainty is represented by the possible action of Ep on early erythroid precursors. In vitro, and in the presence of plateau amounts of BPA, Ep addition is not required for BFU-E–CFU-E differentiation (8, 28a) (some Ep is present, however, in the plated FCS; furthermore, large amounts of BPA may cause in vitro nonphysiological phenomena). On the other hand, Goldwasser (57) suggested that Ep acts at the level of pluripotent stem cell. In line with this concept, BFU-E number and cycling fluctuate in parallel with Ep activity early after erythroid perturbations (26, 27). Macroscopic erythroid bursts, grown in the presence of both Ep and BPA, contain CFU-S (approximately one per burst) (57a): BFU-E may therefore derive from self-replicating CFU-S interacting with BPA and Ep.

In receptor terms, differentiation of erythroid precursors is possibly mediated via a gradual increase of the number of Ep receptors thereon (leading to progression of erythroid commitment and Ep sensitivity), as well as disappearance of CSF receptors. An inverse phenomenon would mediate myeloid differentiation.

In terms of regulating factors, it is generally conceded that BPA modulates proliferation and differentiation of the primitive BFU-E, not only in vitro but also probably in vivo. It is still uncertain, however, whether or not Ep plays an additional regulatory role on early erythroid progenitors. At the level of distal erythroid progenitors, the influence of BPA is progressively dampened, whereas that of Ep becomes gradually more pronounced. In this regard, the possibility exists that Ep operates as a survival factor of differentiating precursors.

Further progress in this crucial area depends on purification of modulating factors, elucidation of cell-cell interactions, and more particularly on in vivo studies, which should confirm and clarify the present concepts derived almost exclusively from in vitro experiments.

Endocrine Influences on Erythropoiesis

In the presence of Ep, various hormones (androgens, GH, thyroid hormones, corticosteroids or their derivatives) enhance the in vitro growth of erythroid precursors (58–62). Thyroid hormones apparently act via a β-adrenergic receptor mechanism (60, 61). These agents also potentiate in vivo Ep production (63). Therefore, they apparently cooperate in the regulation of erythropoiesis, via modulation of both Ep levels and responsiveness to Ep of its target cells. Opposite effects are induced by estrogens (C. Peschle et al, in preparation).

The Hemopoietic Microenvironment: Role of PGE-cAMP Complex and BPA

Physical interaction between hemopoietic cells, possibly via electrochemical or electrotonic influences at the level of cell membranes, has not yet been clearly documented. Thus, it is hypothesized that microenvironmental influences on erythropoiesis are modulated via short-acting chemical mediators.

Kurland et al (64) postulated that CFU-C kinetics, and thus granulopoiesis, may be regulated via macrophagic release of either stimulatory (CSF) or inhibitory (PGE) agents, respectively in the face of low or high macrophage cellularity. Admittedly, PGE exerts its effect via its intracellular mediator, cAMP. A possible role for the PGE-cAMP complex in the regulation of erythropoiesis is under scrutiny. In vitro these agents inhibit cloning of murine BFU-E, and stimulate cloning of human BFU-E, (C. Peschle and G. B. Rossi, in preparation). Their action on either CFU-E growth or heme synthesis is similarly species-dependent; a stimulatory effect has been observed in humans, sheep, and dogs (65, 66), an inhibitory action in rats and guinea pigs (66), and variable influences in mice (66, 67).

Finally it must be emphasized that a major microenvironmental influence on erythropoiesis may be macrophage production of BPA and even perhaps of Ep (see the section on Biogenesis of Ep and BPA).

CLINICAL ASPECTS

Polycythemia Vera

In two females with polycythemia vera (PV), heterozygotes for G-6-PD isoenzymes, normal tissue displayed both B and A enzymes, whereas peripheral erythrocytes, granulocytes, and platelets contained only the type A enzyme (68). Hence PV is apparently a clonal disease of the pluripotent stem cell. Clonal analysis of marrow erythroid precursors[3] in these patients showed that colonies growing in the absence of added Ep contained only the type A enzyme, i.e. derived from the abnormal clone (69). However, addition of Ep induced other erythroid precursors to give rise to colonies containing both A and B enzymes, which suggests that these progenitors did not belong to the PV clone (69). The Ep dependence of erythropoiesis in PV is still uncertain. Ep activity, although markedly less in patients with high hematocrits (70), might still stimulate their erythropoietic rate in view of the apparently enhanced sensitivity to Ep of PV erythroid progenitors (71). Cloning of PV erythroid precursors in Ep-free (serum-free) cultures is clearly necessary to fully elucidate these aspects.

Pure Red Cell Aplasia and Aplastic Anemia

Peripheral lymphocytes from congenital pure red cell aplasia (PRCA) (Blackfan-Diamond syndrome), coincubated with normal marrow, inhibit CFU-E colony formation (72). This inhibitory phenomenon has also been observed in patients with aplastic anemia (73). In both studies, the patients were repeatedly transfused, thus rendering possible, in some cases, a lymphocytic sensitization by the transfused alloantigens. The inhibitory phenomenon has not been observed in other cases of congenital PRCA (74, 75) and untransfused aplastic anemia (76). Similarly, the inhibition by aplastic anemia lymphocytes on normal CFU-C growth (77) has been observed only in a minority of untransfused cases (78)

At the therapeutic level, administration of antilymphocyte globulin (ALG) and/or corticosteroids is beneficial in some patients with aplastic anemia (79, 80). These observations suggest that a T-lymphocyte-suppressing mechanism operates in some patients with aplastic anemia (80). In other cases, intrinsic defects of stem cells and/or hemopoietic precursors, or inhibitory microenvironmental influences, may represent the key pathogenetic mechanism.

In adult PRCA the serum IgG fraction may often inhibit heme synthesis in cultures of both normal and autologous marrow (81, 82); thus the possibility of a transfusion-induced IgG sensitization seems excluded. These patients usually show other immunological abnormalities (thymoma, SLE, myasthenia gravis, etc) (81, 82). Furthermore, immunosuppressive treatment (cyclophosphamide, azathioprine, use of ALG in resistant cases) or removal of the thymoma induce either a temporary or a permanent remission (81, 82). Altogether the autoimmune genesis in some cases of this type of disease seems well documented.

---

[3] CFU-E and BFU-E colonies were not distinguished in these studies.

In the standard type of PRCA (Type I), the IgG inhibits the erythroid marrow, including the erythroblasts (81) and possibly CFU-E and BFU-E pools (82). The IgG may perhaps interact with the receptor for Ep. In a very rare variant of PRCA (Type II), the IgG acted as an antibody neutralizing serum Ep activity, which was nearly absent in spite of low hematocrit values (83). In many cases (Type III), PRCA represents a pre-leukemic condition progressing to acute myeloid leukemia; an IgG inhibitor has not been clearly documented in these patients (82). Finally, the pathogenesis of PRCA associated with chronic lymphocytic leukemia or lymphoma is still unexplored.

Literature Cited

1. Krantz, S. B., Jacobson, L. O., eds. 1970. *Erythropoietin and Regulation of Erythropoiesis*. Univ. Chicago Press, Ill. 329 pp.
2. Gordon, A. S. 1973. Erythropoietin. In *Vitamins Hormones*. 31:105–74
3. Peschle, C. 1975. Regulation of erythropoiesis and its defects. *Br. J. Haematol.* 31 (Suppl.):69–81
4. Metcalf, D. 1977. Hemopoietic colonies. In vitro cloning of normal and leukemic cells. In *Recent Results in Cancer Research*. Berlin: Springer. 227 pp.
5. Murphy, M. J., Peschle, C., Gordon, A. S., Mirand, E. A., eds. 1978. *In Vitro Aspects of Erythropoiesis*. New York: Springer. 280 pp.
6. Golde, D. W., Cline, J. M., Metcalf, D., Fox, C. F., eds. 1978. *Hemopoietic Cell Differentiation*. New York: Academic. 504 pp.
7. Stamatoyannopoulos, G., Nienhuis, A. W., eds. 1979. *Cellular and Molecular Regulation of Hemoglobin Switching*. New York: Grune & Stratton. 792 pp.
8. Iscove, N. N. 1978. Erythropoietin-independent stimulation of early erthropoiesis in adult bone marrow cultures by conditioned media from lectin-stimulated mouse spleen cells. See Ref. 6, pp. 37–52
9. Till, J. E., McCulloch, E. A. 1961. A direct measurement of the radiation sensitivity of normal mouse marrow cells. *Radiat. Res.* 14:213–22
10. Metcalf, D., Johnson, G. R., Mandel, T. E. 1979. Colony formation in agar by multipotential hemopoietic cells. *J. Cell. Physiol.* 98:401–20
11. Metcalf, D., Johnson, G. R. 1979. Interactions between purified GM-CSF, purified erythropoietin and spleen conditioned medium on hemopoietic colony formation in vitro. *J. Cell. Physiol.* 99:159–74
12. Axelrad, A. A., McLeod, D. L., Shreeve, M. M., Heath, D. S. 1974. Properties of cells that produce erythrocytic colonies in vitro. In *Hemopoiesis in Culture*, ed. W. A. Robinson, pp. 226–34. Washington: GPO. 489 pp.
13. Prchal, J. F., Adamson, J. W., Steinman, L., Fialkow, P. J. 1976. The single cell origin of human erythroid colonies. *J. Cell. Physiol.* 89:489–92
14. Wagemaker, G. 1978. Induction of erythropoietin responsiveness in vitro. See Ref. 6, pp. 109–18
15. Lajtha, L. G., Pozzi, L. V., Schofield, R., Fox, M. 1969. Kinetic properties of hemopoietic stem cells. *Cell Tissue Kinet.* 2:39–49
16. Iscove, N. N. 1977. The role of erythropoietin in regulation of population size and cell cycling of early and late erythroid precursors in mouse bone marrow. *Cell Tissue Kinet.* 10:323–34

17. Eaves, C. J., Humphries, R. K., Eaves, A. C. 1979. In vitro characterization of erythroid precursor cells and the eythropoietic differentiation process. See Ref. 7, pp. 251–73
18. Gregory, C. J., Tepperman, A. D., McCulloch, E. A., Till, J. E. 1974. Erythropoietic progenitors capable of colony formation in culture: response of normal and genetically anemic W/W^v mice to manipulations of the erythron. *J. Cell. Physiol.* 84:1–12
19. Hara, H., Ogawa, M. 1977. Erythropoietic precursors in murine blood. *Exp. Hematol.* 5:159–63
20. Nathan, D. G., Chess, L., Hillman, D. G., Clarke, B., Breard, J., Merler, E., Housman, D. E. 1978. Human erythroid burst-forming unit: T-cell requirement for proliferation in vitro. *J. Exp. Med.* 147:324–39
21. Mijake, T., Kung, C. K. H., Goldwasser, E. 1977. Purification of human erythropoietin. *J. Biol. Chem.* 252: 5558–64
22. Filmanowicz, E., Gurney, C. W. 1961. Studies on erythropoiesis. XVI: Responsive to a single dose of erythropoietin in the polycythemic mouse. *J. Lab. Clin. Med.* 57:65–72
23. Goldwasser, E., Inana, G. 1978. Molecular aspects of the initiation of erythropoiesis. See Ref. 6, pp. 15–24
24. Goldwasser, E. 1971. Some molecular aspects of red cell differentiation. In *Regulation of Erythropoiesis,* ed. A. S. Gordon, M. Condorelli, C. Peschle, pp. 227–233 Milano: Ponte. 412 pp.
25. Stholman, F. Jr. 1971. Control mechanisms in erythropoiesis. See Ref. 24, pp. 71–87
26. Peschle, C., Cillo, C., Rappaport, I. A., Magli, M. C., Migliaccio, G., Pizzella, F., Mastroberardino, G. 1979. Early fluctuations of BFU-E pool size after transfusion or erythropoietin treatment. *Exp. Hematol.* 7:87–93
27. Peschle, C., Cillo, C., Migliaccio, G., Lettieri, F. 1980. Fluctuations of BFU-E and CFU-E cycling after erythroid perturbations: correlations with variations of pool size. *Exp. Hematol.* In press
28. Wagemaker, G. 1978. Cellular and soluble factors influencing the differentiation of primitive erythroid progenitor cells (BFU-E) in vitro. See Ref. 5, pp. 44–57

28a. Wagemaker, G. 1979. The erythropoietin-independent stage of *in vitro* erythropoiesis and its relevance to stem cell differentiation. *Exp. Hematol.* 7 (Suppl. No. 6):4

29. Aye, M. T. 1977. Erythroid colony formation in cultures of human marrow: effect of leukocyte conditioned medium. *J. Cell. Physiol.* 91:69–78
30. Fauser, A. A., Messner, H. A. 1978. Granulo-erythropoietic colonies in human bone marrow, peripheral blood and cord blood. *Blood* 52:1243–48
31. Jacobson, L. D., Goldwasser, E., Fried, W., Plzak, L. F. 1957. Role of the kidney in erythropoiesis. *Nature* 179: 633–34
32. Sherwood, J. B., Goldwasser, E. 1978. Extraction of erythropoietin from normal kidneys. *Endocrinology* 103:866–70
33. Erslev, A. J. 1974. In vitro production of erythropoietin by kidneys perfused with a serum-free solution. *Blood* 44:77–85
34. Peschle, C., Condorelli, M. 1975. Biogenesis of erythropoietin: evidence for pro-erythropoietin in a subcellular fraction of kidney. *Science* 190:910–11
35. Gordon, A. S., Cooper, G. W., Zanjani, E. D. 1967. The kidney and erythropoiesis. *Semin. Hematol.* 4:337–58
36. Gordon, A. S., Kaplan, S. M. 1977. Erythrogenin (REF). In *Kidney Hormones,* ed. J. W. Fisher. 2:187–229. London: Academic. 601 pp.
37. Fisher, J. W., Gross, D. M. 1977. Renal prostaglandins and kidney production of erythropoietin. See Ref. 36, pp. 357–85
38. Fisher, J. W., Taylor, G., Porteous, D. D. 1965. Localization of erythropoietin in glomeruli of sheep kidney by fluorescent antibody technique. *Nature* 205: 611–12
39. Zucali, J. R., Stevens, V., Mirand, E. A. 1975. In vitro production of erythropoietin by mouse fetal liver. *Blood* 46:85–90
40. Peschle, C., Marone, G., Genovese, A., Magli, M. C., Condorelli, M. 1976. Erythropoietin production in the rat: additive role of kidney and liver. *Am. J. Physiol.* 230:845–48
41. Fried, W. 1972. The liver as a source of extrarenal erythropoietin production. *Blood* 40:671–77
42. Joyce, R. A., Chervenick, P. A. 1975. Stimulation of granulopoiesis by liver macrophages. *J. Lab. Clin. Med.* 86:112–17
43. Peschle, C., Marone, G., Genovese, A., Rappaport, I. A., Condorelli, M. 1976. Increased erythropoietin production in anephric rats with hyperplasia of the reticuloendothelial system induced by

colloidal carbon or zymosan. *Blood* 47:325–37
44. Anagnostou, A., Schade, S., Barone, J., Fried, W. 1977. Effect of partial hepatectomy on extrarenal erythropoietin production in rats. *Blood* 50:457–62
45. Naughton, B. A., Gordon, A. S., Piliero, S. J., Lin, P. 1978. Extrarenal erythropoietin. See Ref. 5, pp. 194–217
46. Gruber, D. F., Zucali, J. R., Mirand, E. A. 1977. Identification of erythropoietin producing cells in fetal mouse liver cultures. *Exp. Hematol.* 5:392–98
47. Bessis, N. 1958. L'ilot erythroblastique, unitè fonctionnelle de la moëlle osseuse. *Rev. Hematol.* 13:8–13
48. Cline, M. J., Golde, D. W. 1979. Cellular interactions in haematopoiesis. *Nature* 277:177–81
48a. Iscove, N. N., Schreier, M. 1979. Involvement of T cells and macrophages in generation of burst-promoting activity (BPA). *Exp. Hematol.* 7 (Suppl. No. 6):4
49. Cline, M. J., Golde, D. W. 1974. Production of colony-stimulating activity by human lymphocytes. *Nature* 248: 703–4
50. Golde, D. W., Cline, M. J. 1972. Identification of the colony-stimulating cell in human peripheral blood. *J. Clin. Invest.* 51:2981–83
51. Bradley, T. R., Robinson, W., Metcalf, D. 1967. Colony production in vitro by normal, polycythemic and anemic bone marrow. *Nature* 214:511
52. Kubanek, B., Bock, O., Heit, W., Bock, E., Harris, E. B. 1973. Size and proliferation of stem cell compartments in mice after depression of erythropoiesis. In *Hemopoietic Stem Cells. Ciba Found. Symp.* 13:243–55. Amsterdam: North-Holland. 345 pp.
53. Van Zant, G., Goldwasser, E. 1979. Competition between erythropoietic and colony-stimulating factor for target cells in mouse marrow. *Blood* 53: 946–65
54. Crafts, R. C., Meineke, A. A. 1959. The anemia of hypophysectomized animals. *Ann. NY Acad. Sci.* 77:501–17
55. Bol, S., Williams, N. 1980. The maturation state of three types of granulocyte-macrophage progenitor cells from mouse bone marrow. *J. Cell. Physiol.* In press
55a. Burgess, A. W., Camakaris, J., Metcalf, D. 1977. Purification and properties of colony-stimulating factor from mouse lung-conditioned medium. *J. Biol. Chem.* 252:1998–2003
56. Stanley, F. R., Heard, P. M. 1977. Factor regulating macrophage production and growth. Purification and some properties of the colony-stimulating factor from medium conditioned by mouse L cells. *J. Biol. Chem.* 252:4305–12
56a. Fauser, A. A., Messner, H. A. 1979. Identification of megakaryocytes, macrophages and eosinophils in colonies of human bone marrow containing neutrophilic granulocytes and erythroblasts. *Blood* 53:1023–27
57. Goldwasser, E. 1975. Erythropoietin and the differentiation of red blood cells. *Fed. Proc.* 34:2285–92
57a. Humphries, R. K., Jacky, P. B., Dill, F. J., Eaves, A. C. 1979. CFU-S in individual erythroid colonies derived in vitro from adult mouse marrow. *Nature* 279:718–20
58. Singer, J. W., Adamson, J. W. 1976. Steroids and hematopoiesis. III: The response of granulocytic and erythroid colony-forming cells to steroids of different classes. *Blood* 48:855–64
59. Golde, D. W., Bersch, N., Li, C. M. 1977. Growth hormone: species-specific stimulation of erythropoiesis in vitro. *Science* 196:1112–13
60. Nicholas, D., Hoffman, R., Maffei, L. A., Forget, B. G. 1978. Potentiation of human erythropoiesis in vitro by thyroid hormone. *Nature* 272:260–62
61. Popovic, W. J., Brown, J. E., Adamson, J. W. 1977. The influence of thyroid hormones on in vitro erythropoiesis. Mediation by a receptor with beta-adrenergic properties. *J. Clin. Invest.* 60:907–13
62. Golde, D. W., Earsch, M., Cline, M. J. 1976. Potentiation of erythropoiesis in vitro by dexamethasone. *J. Clin. Invest.* 57:57–62
63. Peschle, C., Rappaport, I. A., Magli, M. C., Marone, G., Lettieri, F., Cillo, C., Gordon, A. S. 1978. Role of the hypophysis in erythropoietin production during hypoxia. *Blood* 51:1117–24
64. Kurland, J. J., Broxmeyer, I. A., Pelus, L. M., Bockman, R. S., Moore, M. A. S. 1978. Role for monocyte-macrophage-derived colony-stimulating factor and prostaglandin E in the positive and negative feedback control of myeloid stem cell proliferation. *Blood* 52:388–407
65. Brown, J. E., Adamson, J. W. 1977. Modulation of in vitro erythropoiesis: enhancement of erythroid colony growth by cyclic nucleotides. *Cell Tissue Kinet.* 10:289–98
66. Brown, J. E., Adamson, J. W. 1977. Studies of the influence of cyclic nucleo-

tides on in vitro hemoglobin synthesis. *Br. J. Haematol.* 35:193–208

67. Golde, D. W., Bersch, N., Cline, M. J. 1979. Hormonal effects on erythroid stem cells. See Ref. 7, pp. 305–19
68. Adamson, J. W., Fialkow, P. J., Murphy, S., Prchal, J. F., Steinman, L. 1976. Polycythemia vera: stem cell and probable clonal origin of the disease. *N. Engl. J. Med.* 295:913–16
69. Prchal, J. F., Adamson, J. W., Murphy, S., Steinman, L., Fialkow, P. J. 1978. Polycythemia vera. The in vitro response of normal and abnormal stem cell lines to erythropoietin. *J. Clin. Invest.* 61:1044–47
70. Adamson, J. W. 1968. The erythropoietin/hematocrit relationship in normal and polycythemic man: implications of marrow regulation. *Blood* 32:597–609
71. Zanjani, E. D., Lutton, J. D., Hoffman, R., Wasserman, L. R. 1977. Erythroid colony formation by polycythemia vera bone marrow in vitro. *J. Clin. Invest.* 59:841–48
72. Hoffman, R., Zanjani, E. D., Lutton, J. D., Zalusky, R., Wasserman, L. R. 1976. Suppression of erythroid-colony formation by lymphocytes from patients with aplastic anemia. *N. Engl. J. Med.* 296:10:13
73. Hoffman, R., Zanjani, E. D., Vila, J., Zalusky, R., Lutton, J. D., Waserman, L. R. 1976. Diamond-Blackfan syndrome: lymphocyte-mediated suppression of erythropoiesis. *Science* 193:899–900
74. Nathan, D. G., Clarke, J. B., Hillman, D. G., Alter, B. P., Housman, D. E. 1978. Erythroid precursors in congenital hypoplastic (Diamond-Blackfan) anemia. *J. Clin. Invest.* 61:489–98
75. Freedman, M. H., Saunders, E. F. 1978. Diamond-Blackfan syndrome: evidence against cell-mediated suppression. *Blood* 51:1125–28
76. Adamson, J. W. 1978. Discussion. See Ref. 5, p. 39
77. Kagan, W. A., Ascensao, J. A., Pahwa, R. N., Hansen, J. A., Goldstein, G., Valera, E. B., Inufy, G. S., Moore, M. A. S., Good, R. A. 1976. Aplastic anemia: presence in human marrow of cells that suppress myelopoiesis. *Proc. Natl. Acad. Sci. USA* 73:2890–93
78. Singer, J. W., Brown, J. E., Melody, C. J., Doney, K., Warren, R. P., Storb, R., Thomas, E. D. 1978. Effect of peripheral blood lymphocytes from patients with aplastic anemia on granulocytic colony growth from HLA-matched and mismatched marrows: effect of transfusion sensitization. *Blood* 52:37–46
79. Speck, B., Cornu, P., Niessen, C., Groff, P., Weber, W., Jeannet, M. 1978. On the pathogenesis and treatment of aplastic anemia. In *Experimental Hematology Today*, ed. S. J. Baum, G. D. Ledney, pp. 143–50. New York: Springer. 256 pp.
80. Bagby, G. C. Jr., Goodnight, S. H., Mooney, W. M., Richert-Boe, K. 1979. Prednisone-responsive aplastic anemia: a mechanism of glucocorticoid action. *Blood* 54:322–33
81. Krantz, S. B., Zaentz, S. D. 1977. Pure red cell aplasia. In *The Year in Hematology*, ed. A. S. Gordon, R. Silber, J. Lo Bue, pp. 153–90. New York: Plenum
82. Peschle, C., Marmont, A., Perugini, S., Bernasconi, C., Brunetti, P., Fontana, G., Ghio, R., Resegotti, L., Rizzo, S. C., Condorelli, M. 1976. Physiopathology and therapy of adult pure red cell aplasia (PRCA): a cooperative study. In *Aplastic Anemia*, ed. S. Hibino, F. Takata, N. T. Shahidi, pp. 285–303. Univ. Tokyo Press. 443 pp.
83. Peschle, C., Marmont, A. M., Marone, G., Genovese, A., Sasso, G. F., Condorelli, M. 1975. Pure red cell aplasia: studies on an IgG serum inhibitor neutralizing erythropoietin. *Br. J. Haematol.* 30:411–17

Ann. Rev. Med. 1980. 31:315–27
Copyright © 1980 by Annual Reviews Inc. All rights reserved

VASOPRESSIN FUNCTION IN THE SYNDROME OF INAPPROPRIATE ANTIDIURESIS

❖7363

R. Zerbe, M.D., L. Stropes, M.D., and G. Robertson, M.D.

Department of Endocrinology, Indiana University and Veterans Administration Medical Center, Indianapolis, Indiana 46202

INTRODUCTION

In 1938 Winkler & Crankshaw (1) described a group of patients who had hypochloremia and excessive urinary excretion of chloride despite apparently normal kidneys and adrenal glands at autopsy. Subsequently other investigators reported patients with hyponatremia and urinary sodium wasting (2–5). In 1957, Schwartz et al (6) described similar findings in association with impaired urinary dilution in two patients with bronchogenic carcinoma. They concluded from indirect evidence that the syndrome was due to inappropriate secretion of the antidiuretic hormone, arginine vasopressin. It is now recognized that the syndrome of inappropriate antidiuresis (SIAD) occurs in a great variety of malignant and nonmalignant diseases, most commonly those involving the central nervous system and lungs (Table 1) (7–51).

The pathophysiology of this syndrome has been well delineated and reviewed elsewhere (7, 8). Briefly, Schwartz and Bartter originally deduced from earlier studies of the effect of exogenous antidiuretic hormone in normal humans (52) that the syndrome resulted from inappropriate secretion of antidiuretic hormone in the face of excessive water intake. This constellation of defects results in hyponatremia due to dilution and excessive urinary excretion of sodium caused by the water retention. Hence, the cardinal features of SIAD as defined by Bartter & Schwartz (7) are (*a*) hyponatremia with corresponding hypoosmolality of the plasma, (*b*) osmolality of the urine greater than that appropriate for concomitant plasma

0066-4219/80/0401-0315$01.00

Table 1 Disorders associated with SIADH

1. Malignant tumors
    - A. Carcinoma of the lung
    - B. Carcinoma of the duodenum
    - C. Carcinoma of the pancreas
    - D. Thymoma
    - E. Mesothelioma
    - F. Carcinoma of the bladder
    - G. Carcinoma of the ureter
    - H. Prostatic carcinoma
    - I. Lymphoma
    - J. Ewing's sarcoma
2. Disorders of the central nervous system
    - A. Meningitis
    - B. Head trauma
    - C. Brain abscess
    - D. Brain tumors
    - E. Encephalitis
    - F. Guillain-Barré syndrome
    - G. Acute intermittent porphyria
    - H. Subarachnoid hemorrhage
    - I. Cerebellar and cerebral atrophy
    - J. Cavernous sinus thrombosis
    - K. Neonatal hypoxia
    - L. Hydrocephalus
    - M. Shy-Drager syndrome
    - N. Rocky Mountain spotted fever
    - O. Delirium tremens
3. Diseases of the lung
    - A. Pneumonia
    - B. Tuberculosis
    - C. Cavitation
    - D. Empyema
    - E. Cystic fibrosis
    - F. Pneumothorax
    - G. Asthma
    - H. Positive pressure breathing
4. Drugs
    - A. Vasopressin
    - B. Oxytocin
    - C. Vincristine, Vinblastine
    - D. Chlorpropamide
    - E. Thiazide diuretics
    - F. Phenothiazines
    - G. Monoamine oxidose inhibitors
    - H. Carbamazepine
    - I. Clofibrate
    - J. Nicotine
5. Acute psychosis
6. Postoperative period
7. Endocrine diseases
    - A. Myxedema
8. "Idiopathic"

osmolality, (*c*) excessive renal excretion of sodium, (*d*) absence of clinical evidence of edema-forming states or of volume depletion, (*e*) normal renal function, and (*f*) normal adrenal function. In this review of recent developments in the pathogenesis and treatment of SIAD we have limited ourselves to patients who meet these criteria.

VASOPRESSIN FUNCTION

General Studies

Only within the last decade have assay techniques for plasma vasopressin been sufficiently sensitive and specific to characterize the hormone within the physiologic range. Such studies have shown that plasma vasopressin levels in normal subjects are highly dependent on plasma osmolality (53). Above a level of about 280 mOsmol/kg, plasma vasopressin increases linearly with increasing tonicity. Below this threshold level, vasopressin is uniformly suppressed to less than 1 pg/ml (54), a concentration low enough to permit a diuresis of 10–20 liters per day. Since this rate of output is sufficient to offset all but the most excessive rates of water intake, the ability to maximally suppress vasopressin secretion provides an important defense mechanism against dilutional hyponatremia.

Measurements of plasma or urinary vasopressin have abundantly confirmed that the hormone is inadequately suppressed in most patients with SIAD (37, 55–63). Our series of 79 patients, the largest yet reported, is fully consistent with this finding (Figure 1). It should be noted, however, that the vasopressin levels of over 80% of the patients in our study were comparable to those seen in normally hydrated healthy adults (Figure 1). Hence, they can be recognized as inappropriate only in relation to the hypotonicity of body fluids.

The abnormal osmoregulation of vasopressin in patients with SIAD is further characterized by sequentially measuring the hormone during hypertonic saline and/or water loading. When performed in normal subjects,

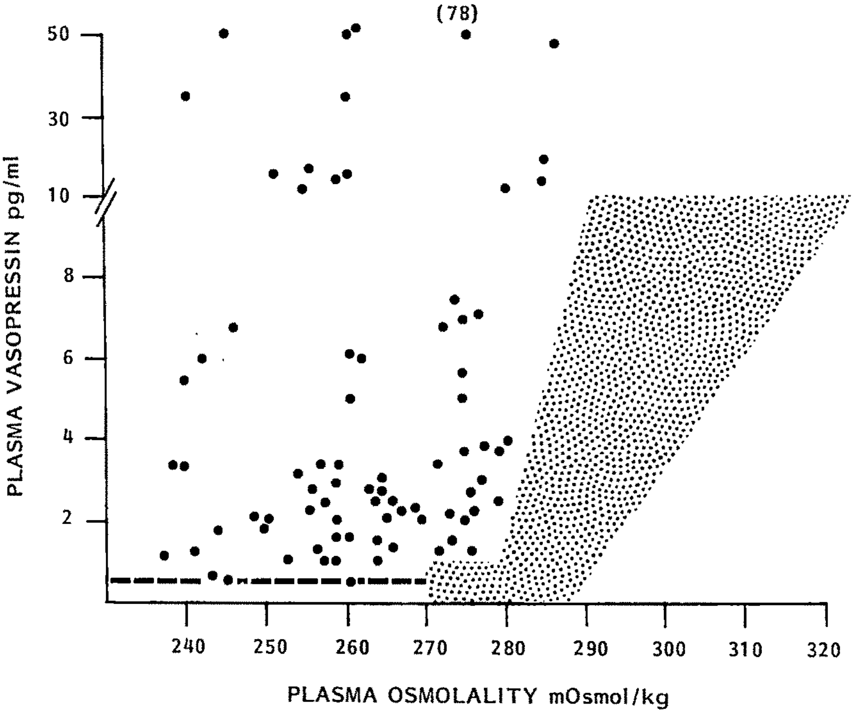


Figure 1 Relationship between plasma vasopressin and osmolality in 79 patients with clinical features of SIAD. Shaded area represents the range of normal values.

these tests produce closely related changes in vasopressin and osmolality that can be described in terms of a linear regression function. With such an analysis, the correlation coefficient or *r* value reflects the precision of the response, the slope of line reflects its sensitivity, and the *x*-intercept provides a measure of the threshold or set of the system. In normal subjects, the results obtained are precise and reproducible on repeat study (Zerbe and Robertson, unpublished observations). When 43 patients with clinical features of SIAD were analyzed in this way, four types of vasopressin response were noted.

The first and most aberrant of these defects occurred in 37% of our patients. It was characterized by wide erratic fluctuations of vasopressin which appeared to be completely independent of osmotic control (Figure 2*a*). Though the precise mechanism is not known, this erratic release could result from random ectopic release of vasopressin, or from variation in posterior pituitary secretion due to rapid fluctuation of nonosmotic stimuli. The high incidence of this type of abnormality points out the need for numerous determinations of vasopressin when assessing the efficacy of a treatment for SIAD, since wide random variation may occur independently of the therapeutic interventions.

A second group of patients comprising 33% of the total demonstrated release of vasopressin which, as in normal subjects, correlated highly with plasma osmolality (Figure 2*b*). In these patients, however, the *x*-intercept (or osmotic threshold) of the regression line was abnormally low. Therefore, the osmoregulation of vasopressin appears to have been preserved but initiation of release was premature. Such resetting of the osmostat could occur by at least three mechanisms. First, hypovolemia and/or hypotension leads to resetting (64), but it cannot be responsible in these patients since by definition they are clinically normovolemic and normotensive. Second, resetting may occur as a result of interruption of afferent baroregulatory pathways by such processes as carcinomatous invasion or autonomic neuropathy. The interruption of these normally inhibitory pathways might lead to resetting by providing false signals of volume depletion and/or hypotension (65). Third, since stimulation of the vasopressin osmoreceptor is probably mediated by an osmotically induced dehydration of the neuron, loss of solute from the osmoreceptor could also cause resetting. Loss of intracellular solutes from the osmoreceptor might be a primary defect such as that hypothesized for the "sick cell syndrome" (66), or it could be a secondary adaptation to chronic hyponatremia (67). Whatever the pathogenic mechanism, it should be noted that patients with resetting of the osmostat behave differently from those with other forms of SIAD since they can produce maximally dilute urine if made sufficiently hyponatremic (68, 69).

In 16% of our patients there appeared to be a constant nonsuppressible "leak" of vasopressin with otherwise normal osmoreceptor function. Within the physiologic range of plasma osmolality, vasopressin release stimulated appropriately, but below 278 mOsmol/kg, the hormone stabilized at a constant, inappropriately high level which could not be suppressed by hypotonicity (Figure 2*c*). Such a defect might result from constant ectopic or neurohypophyseal release. Nonsuppressible central secretion of vasopressin may be due to hyperfunction of stimulatory neurons, hypofunction of proposed inhibitory neurons (65, 70), or injury to the neurohypophysis itself.

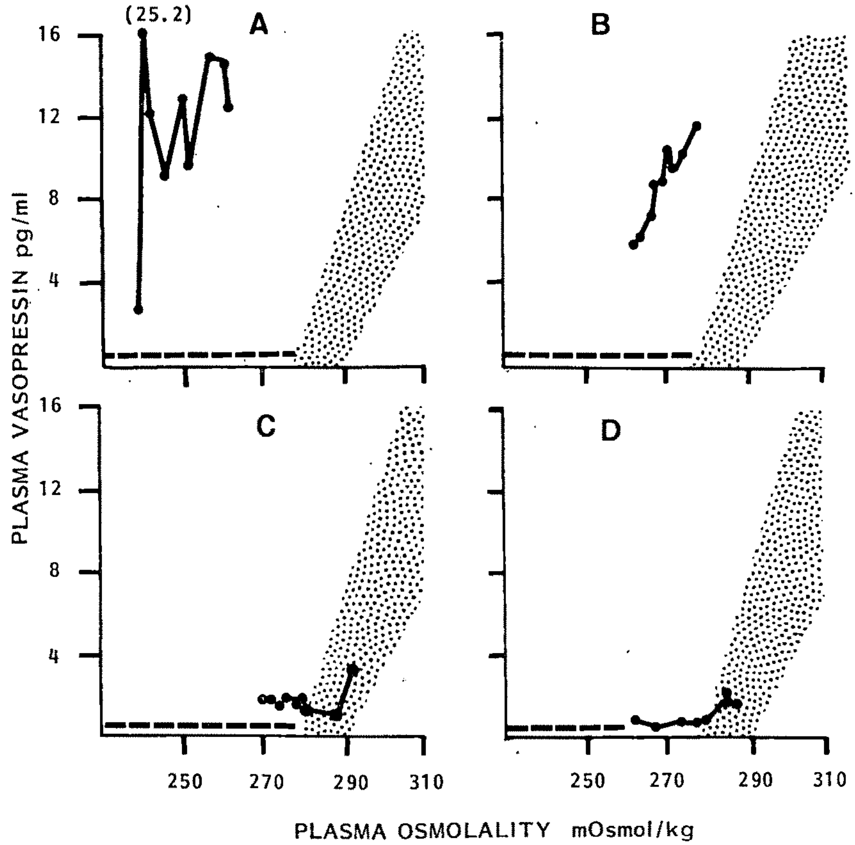


Figure 2 Examples of the four patterns of vasopressin release observed in patients with SIAD during osmotic challenge: (*a*) erratic release, (*b*) reset osmostat, (*c*) vasopressin leak, (*d*) hypovasopressinemic antidiuresis.

The last and smallest group, comprising 14% of our patients, exhibited what we have termed "hypovasopressinemic antidiureses." They exhibited no detectable abnormalities in vasopressin secretion, yet they could neither maximally dilute their urine nor excrete a water load normally. This type of abnormality in water balance could be due to either increased renal sensitivity to arginine vasopressin or to the presence of another antidiuretic substance immunologically distinct from this hormone. Increased sensitivity to vasopressin might occur as a result of changes in the affinity or number of renal vasopressin receptors, the regulation of which is not understood. It could also be mediated by a deficiency of renal prostaglandins since these substances appear to antagonize vasopressin action in vivo (71) as well as in vitro (72). The presence of an immunologically distinct antidiuretic substance such as arginine vasotocin might also explain antidiuresis in the absence of detectable vasopressin levels. Vasotocin has been identified in both human fetal pineal glands (73) and adult cerebrospinal fluid (74) and its antidiuretic effect in mammals is equal to that of vasopressin (Zerbe and Stropes, unpublished observations). Release of this ancestral hormone could result from some defect in the blood-CSF barrier or from synthesis in malignant tissue that has dedifferentiated to a more primitive metabolic state. Whatever the pathogenesis, however, the apparent existence of this hypovasopressinemic form suggests that the term inappropriate antidiuretic hormone (SIADH) ought to be reserved for those patients in whom vasopressin is shown to be elevated by direct assay. In others, it would be more accurate to speak of the syndrome of inappropriate antidiuresis (SIAD).

With reference to this system of classification, three additional points need to be made. First, because SIAD is often transient, repeat studies are difficult to obtain. In the few patients studied more than once, the type of vasopressin defect has been reproducible but further studies are needed to verify this observation. Second, more than one type of defect may coexist since both hypovasopressinemic antidiuresis and resetting of the osmostat have been observed in the same patient (75). The reason for this association is opened to speculation, but may indicate that the two defects are causally interrelated at least in some patients (see above). Third, none of the four types of osmoregulatory defects was found to be associated with any specific underlying illness. This lack of association could be related to any of three possibilities.

First, the underlying illness may be coincidental rather than causal. This seems unlikely considering the frequency with which certain diseases are associated with SIAD. Second, it is possible that a certain underlying illness can cause SIAD by more than one mechanism. This appears to be true for cancer and may apply to other diseases as well. Finally, different diseases may produce the syndrome by a common final pathway. For example, resetting of the osmostat might result from loss of baroregulatory input or loss of osmoreceptor solute as previously noted. Further study of these patients and the various factors involved in the control of vasopressin release are needed to distinguish among these various possibilities.

Cancer

Inappropriate antidiuresis has been associated with a variety of malignant tumors, particularly oat cell carcinoma (Table 1). Schwartz and Bartter originally proposed a nonectopic mechanism for inappropriate secretion of antidiuretic hormone in these patients (6). The possibility of ectopic production of the hormone by malignant tumors was first suggested in the early 1960s when a bronchogenic carcinoma removed from a patient with features of SIAD was shown to contain large amounts of antidiuretic activity (76). This conclusion was confirmed when a similar tumor was found and shown to synthesize vasopressin in vitro (77). This observation lead to the widespread belief that all cases of SIAD associated with malignancies were due to ectopic production. However, two lines of evidence challenge this view. In the first place, only four of nine such tumors studied subsequently demonstrated detectable amounts of antidiuretic activity (78). Second, many patients with cancer and inappropriate antidiuresis have resetting of the osmostat. It is highly unlikely that malignant tissue would acquire through dedifferentiation the capacity to both synthesize and osmoregulate vasopressin since these two functions are normally subserved by different neurons. A more plausible explanation is that many tumors act by resetting the osmostat or otherwise causing the neurohypophysis to release vasopressin inappropriately. Such effects could be mediated by: (*a*) hypovolemia and/or hypotension, (*b*) obstruction of the vena cava (effective hypovolemia), (*c*) invasion of the vagus nerve (baroreceptor denervation), (*d*) metastases in the hypothalamus or other regulatory centers, (*e*) carcinomatous neuropathy (peripheral and/or central), and (*f*) ectopic production of a neurotransmitter or a releasing factor for vasopressin (79).

Hypothyroidism

Impaired free water excretion and hyponatremia in the presence of severe hypothyroidism has been recognized for 25 years (80) but the incidence and etiology of this disorder remain controversial. Early reports demonstrated impaired water excretion in all of eight myxedematous patients studied (80). However, subsequent studies found a lower incidence (81, 82), and a small group of hypothyroid patients that we studied showed no abnormality in water excretion. These findings suggest that inappropriate antidiuresis may not be caused by hypothyroidism per se, but by some other variable related to the duration, severity, or complications of the thyroid deficiency. When present, moreover, there is still uncertainty about the cause of the impaired water excretion since vasopressin is often (82), but not always (83), inappropriately high.

TREATMENT

The clinical consequences of inappropriate antidiuresis are varied. They are due largely to the cerebral edema that results from hypotonicity of body fluids and can range from mild anorexia and confusion to stupor, convulsions, and coma (8). Both the degree and rate of development of hypoosmolality are probably important determinants of the neurologic symptoms because there is some evidence that the brain can counteract cerebral edema by gradually reducing intracellular solutes (84).

Treatment of SIAD will vary depending on the clinical situation. The traditional method, fluid restriction, is safe, rational, and effective. However, correction of the hyponatremia is slow and difficult to maintain chronically. Hypertonic saline is useful in emergent situations associated with coma or convulsions, but again, this approach is not suitable for long-term use.

In recent years other forms of therapy that may be useful in the treatment of SIAD have become available. They involve the use of drugs that either suppress the release of vasopressin or inhibit its action on the kidney. Ethanol is known to inhibit the release of vasopressin (85, 86) and has been proposed as therapy for SIAD (51). However, this approach is rarely practical because the effects of ethanol are short-lived, inconsistent, and require intoxicating levels (86). Diphenylhydantoin was reported initially to be effective only when given intravenously to patients with nonmalignant forms of SIAD (55). More recently, oral diphenylhydantoin was successfully used to treat a patient with SIAD (87). Oxilorphan, butorphanol (88), and other opiates (89) suppress vasopressin release in rats; the first has also been reported to increase water excretion in one patient with SIAD (90). Whether oral diphenylhydantoin and/or opiates prove to be effective in other such patients remains to be established. Given the heterogeneity of the vasopressin dysfunction in patients with SIAD it would be surprising if any drug that suppresses vasopressin release proves to be uniformly effective.

The most practical agents currently available for treatment of chronic SIAD are those that inhibit the antidiuretic effects of vasopressin. Lithium causes nephrogenic diabetes insipidus in about 20% of people with normal antidiuretic function (91) and has been used to correct hyponatremia in patients with inappropriate antidiuresis (92). However, it is effective in only a minority of patients, and its several toxic side effects including sluggishness, tremor, gastrointestinal disturbances, cardiotoxicity, and thyroid dysfunction severely limit its use (93–95).

Demeclocycline appears to be a safer, more effective method of inducing nephrogenic diabetes insipidus. It is an antibiotic of the tetracycline group and inhibits the action of vasopressin apparently by impairing both the generation and action of cyclic-adenosine monophosphate (96–98). Demeclocycline is relatively safe, but potential adverse effects include nausea, photosensitivity, azotemia, and superinfection with resistant bacteria. At usual doses of 600–1200 mg/day it produces reversible nephrogenic diabetes insipidus in virtually all patients with inappropriate antidiuresis and currently is the preferred form of therapy in patients with chronic SIAD (93, 99).

CONCLUSION

SIAD is associated with a wide variety of diseases and medications. In most cases it arises from inadequately suppressed levels of plasma vasopressin, which can result from any one of at least three different types of osmoregulatory defect. A few patients have hypovasopressinemic antidiuresis that is characterized by low free water clearance in the absence of detectable vasopressin. These four types of diluting defects are not specific for any disease process, and the responsible pathogenetic mechanisms are open to speculation. The most effective treatment of chronic SIAD appears to be demeclocycline, which acts by producing a state of nephrogenic diabetes insipidus. Other medications suppress the central release of vasopressin but have not been shown to be uniformly effective.

Literature Cited

1. Winkler, A. W., Crankshaw, O. F. 1938. Chloride depletion in conditions other than Addison's disease. *J. Clin. Invest.* 17:1–6
2. Sims, E. A. H., Welt, L. G., Orloff, J., Needham, J. W. 1950. Asymptomatic hyponatremia in pulmonary tuberculosis. *J. Clin. Invest.* 29:1545–57
3. Rapoport, S., West, C. D., Brodsky, W. A. 1951. Salt losing conditions: the renal defect in tuberculosis meningitis. *J. Lab. Clin. Med.* 37:550–61
4. Harrison, H. E., Finberg, L., Fleishman, E. 1951. Disturbances of ionic equilibrium of intracellular and extracellular electrolytes in patients with tuberculosis meningitis. *J. Clin. Invest.* 31:300–8
5. Cheek, D. B. 1956. Further observations on electrolyte change in tuberculous meningitis. The ratio of the concentrations of bromide in serum and cerebrospinal fluid. *Pediatrics* 18:218–26
6. Schwartz, W. B., Bennett, W., Curelop, S., Bartter, F. C. 1957. A syndrome of renal sodium loss and hyponatremia resulting from inappropriate secretion of antidiuretic hormone. *Am. J. Med.* 23:529–42
7. Bartter, F. C., Schwartz, W. B. 1967. The syndrome of inappropriate secretion of antidiuretic hormone. *Am. J. Med.* 42:790–806
8. Bartter, F. C. 1973. The syndrome of inappropriate antidiuretic hormone (SIADH). *Dis. Month* Nov:1–47

9. Perks, W. H., Crow, J. C., Green, M. 1978. Mesothelioma associated with the syndrome of inappropriate secretion of antidiuretic hormone. *Am. Rev. Respir. Dis.* 117:789–94

10. Kaye, S. B., Ross, E. J. 1977. Inappropriate antidiuretic hormone secretion in association with carcinoma of the bladder. *Postgrad. Med. J.* 53:274–77

11. Coscia, M., Brown, H. D., Miller, M., Tanaka, K., Nicholson, W. E., Parks, K. R., Orth, D. N. 1977. Ectopic production of antidiuretic hormone, adrenocorticotrophic hormone and beta-melanocyte stimulating hormone by an oat cell carcinoma of the lung. *Am. J. Med.* 62:303–7

12. Hirata, Y., Matsukura, S., Imura, H., Yakura, T., Ihjima, S. 1976. Two cases of multiple hormone-producing small cell carcinoma of the lung: coexistence of tumor ADH, ACTH, and beta-MSH. *Cancer* 38:2575–82

13. Sacks, S. A., Rhodes, D. B., Malkasian, D. R., Rosenbloom, A. A. 1975. Prostatic carcinoma producing syndrome of inappropriate secretion of antidiuretic hormone. *Urology* 6:489–92

14. Hamilton, D. V. 1978. Inappropriate secretion of antidiuretic hormone associated with cerebellar and cerebral atrophy. *Postgrad. Med. J.* 54:427–28

15. Feinfeld, D. A., Al-Achkar, G., Lipner, H. I., Chiryil, S. J., Hakim, J., Avram, M. M. 1978. Syndrome of inappropriate secretion of antidiuretic hormone: association with cavernous sinus thrombosis. *J. Am. Med. Assoc.* 240:856–57

16. Moylan, F. M., Herrin, J. T., Krishnamoorthy, K., Todres, I. D., Shannon, D. C. 1978. Inappropriate antidiuretic hormone secretion in premature infants with cerebral injury. *Am. J. Dis. Child.* 132:399–402

17. Kaplan, S. L., Feigin, R. D. 1978. The syndrome of inappropriate secretion of antidiuretic hormone in children with bacterial meningitis. *J. Pediatr.* 92: 758–61

18. Kaplan, S. L., Feigin, R. D. 1978. Inappropriate secretion of antidiuretic hormone complicating neonatal hypoxic-ischemic encephalopathy. *J. Pediatr.* 92:431–33

19. Ganfin, L., Skowsky, W. R., Goodman, S. L. 1977. Release of antidiuretic hormone during mass-induced elevation of intracranial pressure. *J. Neurosurg.* 46:627–37

20. Davis, B. P., Matrikas, V. J. 1976. Inappropriate secretion of antidiuretic hormone after cerebral injury. *J. Oral Surg.* 34:609–15

21 Imbean, S. A., Rock, W. 1976. Syndrome of inappropriate antidiuretic hormone secretion with subarachnoid hemorrhage. *Wis. Med. J.* 75:825–28

22. Peterson, D. T., Marshall, W. H. 1975. Polydipsia and inappropriate secretion of antidiuretic hormone associated with hydrocephalus. *Ann. Intern. Med.* 83:675–76

23. Lockwood, A. H. 1976. Shy-Drager syndrome with abnormal respirations and antidiuretic hormone release. *Arch. Neurol.* 33:292–95

24. Sexton, D. J., Clapp, J. 1977. Inappropriate antidiuretic hormone secretion: occurrence in a patient with Rocky Mountain Spotted Fever. *Arch. Intern. Med.* 137:362–63

25. Mor, J., Ben-Galin, E., Abrahamov, A. 1975. Inappropriate antidiuretic hormone secretion in an infant with severe pneumonia. *Am. J. Dis. Child.* 129: 133–35

26. Davies, M. G. 1977. Pneumonia causing the syndrome of inappropriate secretion of antidiuretic hormone. *Br. J. Clin. Pract.* 31:82–84

27. Charles, R., Rees, J. R. 1975. Inappropriate secretion of antidiuretic hormone in pneumonia. *Postgrad. Med. J.* 51:633–34

28. Pollard, R. B. 1975. Inappropriate secretion of antidiuretic hormone associated with adenovirus pneumonia. *Chest* 68:589–91

29. Sahn, S. A., Skeff, K. M. 1977. Tuberculous pneumonia with the syndrome of inappropriate secretion of antidiuretic hormone: cause of adult respiratory distress syndrome. *Chest* 72:678–80

30. Gerdman, J. C. 1977. Putrid pulmonary abscess and empyema with inappropriate secretion of antidiuretic hormone. *Chest* 72:796–99

31. Petty, B. G. 1977. Inappropriate secretion of ADH in empyema. *Am. Rev. Respir. Dis.* 116:791–92

32. Petty, B. G., Smith, C. R. L. 1977. The syndrome of inappropriate secretion of antidiuretic hormone associated with anaerobic empyema. *Am. Rev. Respir. Dis.* 115:685–88

33. Schiffer, M., Lynch, R. 1977. Inappropriate secretion of ADH as a cause of hyponatremia in cystic fibrosis. *J. Pediatr.* 91:850–51

34. Cohen, L. F., DiSant-Agnese, P. A., Taylor, A., Gill, J. R. Jr. 1977. The syndrome of inappropriate ADH as a cause

of hyponatremia in cystic fibrosis. *J. Pediatr.* 90:574–78
35. Paxson, C. L. Jr., Stoerner, J. W., Denson, S. E., Adcock, E. W. III, Morriss, F. H. Jr. 1977. Syndrome of inappropriate antidiuretic hormone secretion in neonates with pneumothorax or atelectasis. *J. Pediatr.* 91:459–63
36. Crigler, J. F. Jr. 1976. Commentary: on the use of pitressin in infants with neurogenic diabetes insipidus. *J. Pediatr.* 88:295–96
37. Robertson, G. L., Bhoopalam, N., Zelkowitz, L. J. 1973. Vincristine neurotoxicity and abnormal secretion of antidiuretic hormone. *Arch. Intern. Med.* 132:717–20
38. Stuart, M. J., Cuaso, C., Miller, M., Oski, F. A. 1975. Syndrome of recurrent increased secretion of antidiuretic hormone following multiple doses of vincristine. *Blood* 45:315–20
39. Hamilton, D. V. 1978. Chlorpropamide-induced inappropriate secretion of antidiuretic hormone. *Practitioner* 220:469–71
40. Luboshitzky, R., Talor, Z., Barzilar, D. 1978. Chlorthalidone-induced syndrome of inappropriate secretion of antidiuretic hormone. *J. Clin. Pharmacol.* 18:336–39
41. Peterson, J. C., Pollack, R. W., Mahoney, J. J., Fuller, T. J. 1978. Inappropriate antidiuretic hormone secondary to monoamine oxidase inhibition. *J. Am. Med. Assoc.* 239:1422–23
42. Matak, F., Kalyanaraman, K. 1977. Syndrome of inappropriate secretion of antidiuretic hormone in patients treated with psychotherapeutic drugs. *Arch. Neurol.* 34:374–75
43. Zuniga, J. S., Hazan, T. 1975. Fluphenazine and inappropriate secretion of antidiuretic hormone. *Ann. Intern. Med.* 83:735–37
44. Sordillo, P., Sagransky, D. M., Mercado, R. M., Michelis, M. F. 1978. Carbamazepine-induced syndrome of inappropriate antidiuretic hormone secretion: reversal by concomitant phenytoin therapy. *Arch. Intern. Med.* 138:299–301
45. Rado, J. P., Juhos, E., Sawinsky, I. 1975. Dose-response relations in drug-induced inappropriate secretion of ADH: effects of clofibrate and carbamazepine. *J. Clin. Pharmacol. Biopharm.* 12:315–19
46. Winter, S. C., Arbus, G. S. 1977. Syndrome of inappropriate secretion of antidiuretic hormone secondary to vinblastine overdose. *Can. Med. Assoc. J.* 117:1134
47. Ginsberg, S. J., Comis, R. L., Fitzpatrick, A. V. 1977. Vinblastine and inappropriate ADH secretion. *N. Engl. J. Med.* 296:941
48. Dubovsky, S. L., Grabon, S., Berl, T., Schrier, R. W. 1973. Syndrome of inappropriate secretion of antidiuretic hormone with exacerbated psychosis. *Ann. Intern. Med.* 79:551–54
49. Raskind, M. A., Orenstein, H., Christopher, T. G. 1975. Acute psychosis, increased water ingestion and inappropriate antidiuretic hormone secretion. *Am. J. Psychiatry* 132:907–10
50. Fowler, R. C., Kranfol, Z. A., Perry, P. J. 1977. Water intoxication psychosis and inappropriate secretion of antidiuretic hormone. *Arch. Gen. Psychiatry* 34:1097–99
51. Mendoza, S. 1976. Syndrome of inappropriate antidiuretic hormone secretion. *Pediatr. Clin. North Am.* 23: 681–90
52. Leaf, A., Bartter, F. C., Santos, R. F., Wrong, O. 1952. Evidence in man that urinary electrolyte loss induced by pitressin is a function of water retention. *J. Clin. Invest.* 32:868–78
53. Robertson, G. L., Mahr, E. A., Athar, S., Sinha, T. 1973. Development and clinical application of a new method for the radioimmunoassay of arginine vasopressin in human plasma. *J. Clin. Invest.* 52:2340–52
54. Robertson, G. L., Shelton, R. L., Athar, S. 1976. The osmoregulation of vasopressin. *Kidney Int.* 10:25–37
55. Fichman, M. P., Bethune, J. E. 1968. The role of adrenocorticoids in the inappropriate antidiuretic hormone syndrome. *Ann. Intern. Med.* 68:806–20
56. Bauman, G., Lopez-Amor, E., Dingman, J. F. 1972. Plasma arginine vasopressin in the syndrome of inappropriate antidiuretic hormone secretion. *Am. J. Med.* 52:19–24
57. Shimamoto, K., Murase, T., Yamaji, T. 1976. A heterologous radioimmunoassay for arginine vasopressin. *J. Lab. Clin. Med.* 87:338–44
58. Morton, J. J., Padfield, P. L., Forsling, M. L. 1975. A radioimmunoassay for plasma arginine-vasopressin in man and dogs: application to physiological and pathological states. *J. Endocrinol.* 65:411–24
59. Miller, M., Moses, A. M. 1972. Urinary antidiuretic hormone in polyuric disorders and in inappropriate ADH syndrome. *Ann. Intern. Med.* 77:715–21

60. Merkelbach, U., Czernichow, R. C., Gaillard, R. C., Vallotton, M. B. 1975. Radioimmunoassay of (8-arginine)-vasopressin. *Acta Endocrinol.* 80: 453–64

61. Beardwell, C. G. 1971. Radioimmunoassay of arginine vasopressin in human plasma. *J. Clin. Endocrinol. Metab.* 33:254–60

62. DeFronzo, R. A., Colvin, O. M., Braine, H., Robertson, G. L., Davis, P. J. 1974. Cyclophosphamide and the kidney. *Cancer* 33:483–91

63. Zimbler, H., Robertson, G. L., Bartter, F. C., Delea, C. S., Pomeroy, T. 1975. Ewings sarcoma as a cause of the syndrome of inappropriate secretion of antidiuretic hormone. *J. Clin. Endocrinol. Metab.* 41:390–91

64. Robertson, G. L., Athar, S. 1976. The interaction of blood osmolality and blood volume in regulating plasma vasopressin in man. *J. Clin. Endocrinol. Metab.* 42:613–20

65. Robertson, G. L. 1979. Control of the posterior pituitary and antidiuretic hormone secretion. In *Contributions to Nephrology,* ed. J. Brod, Basel: Karger. In press

66. Flear, D. T. G., Singh, C. M. 1973. Hyponatremia in sick cells. *Br. J. Anesth.* 45:976–94

67. Grantham, J. J. 1977. Pathophysiology of hyposmolar conditions: a cellular perspective. In *Disturbances in Body Fluid Osmolality,* ed. T. E. Andreoli, J. J. Grantham, F. C. Rector Jr., 10:217–25. Bethesda, Md: Am. Physiol. Soc. 349 pp.

68. Michelis, M. F., Fusco, R. O., Bragdon, R. W., Davis, B. D. 1974. Reset of osmoreceptors in association with normovolemic hyponatremia. *Am. J. Med. Sci.* 267:267–73

69. DeFronzo, R. A., Goldberg, M., Agus, Z. S. 1976. Normal diluting capacity in hyponatremic patients: reset osmostat or a variant of SIADH. *Ann. Intern. Med.* 84:538–42

70. Vincent, J. D., Arnauld, E., Bioulac, B. 1972. Activity of osmosensitive single cells in the hypothalamus of the behaving monkey during drinking. *Brain Res.* 44:371–84

71. Anderson, R. J., Berl, T., McDonald, K. M., Schrier, R. W. 1975. Evidence for an *in vivo* antagonism between vasopressin and prostaglandin in the mammalian kidney. *J. Clin. Invest.* 56:420–26

72. Orloff, J., Handler, J. S., Bergstrom, S. 1965. Effect of prostaglandin (PGE_1) on the permeability response of toad bladder to vasopressin, theophylline and adenosine 3',5'-monophosphate. *Nature* 205:397–98

73 Legros, J. J., Louis, F., Demoulin, A., Franchimont, P. 1976. Immunoreactive neurophysin and vastocin in human foetal pineal glands. *J. Endocrinol.* 68: 289–90

74. Pavel, S. 1970. Tentative identification of arginine vasotocin in human cerebrospinal fluid. *J. Clin. Endocrinol.* 31:369–71

75. Baylis, P. H., Robertson, G. L. 1979. Absent baroregulation of vasopressin with hypertension and inappropriate antidiuresis. *Clin. Res.* 27:447A

76. Amatruda, T. T. Jr., Mulrow, P. J., Gallagher, J. C., Sawyer, W. H. 1963. Carcinoma of the lung with inappropriate antidiuresis. *N. Engl. J. Med.* 269:544–49

77. George, J. M., Capen, C. C., Phillips, A. S. 1972. Biosynthesis of vasopressin *in vitro* and ultrastructure of a bronchogenic carcinoma: patient with syndrome of inappropriate secretion of antidiuretic hormone. *J. Clin. Invest.* 51: 141–48

78. Vorherr, H., Massry, S. G., Utiger, R. D., Kleeman, C. R. 1968. Antidiuretic principle in malignant tumor extracts from patients with inappropriate ADH syndrome. *J. Clin. Endocrinol. Metab.* 28:162–68

79. Robertson, G. L. 1978. Cancer and inappropriate antidiuresis. In *Biological Markers of Neoplasia: Basic and Applied Aspects,* ed. R. W. Ruddon, pp. 277–93. New York: Elsevier North-Holland. 590 pp.

80. Crispell, K. R., Parson, W., Sprinkle, P. 1954. A cortisone-resistant abnormality in the diuretic response to ingested water in primary myxedema. *J. Clin. Endocrinol. Metab.* 14:640–44

81. Vogt, J. H. 1960. Impaired water excretion capacity in primary myxoedema improved by corticosteroids, corticotrophin and thyroid substitution. *Acta Endocrinol.* 35:277–92

82. Skowsky, W. R., Kikuchi, T. A. 1978. The role of vasopressin in the impaired water excretion of myxedema. *Am. J. Med.* 64:613–21

83. Macaron, C., Famuyiwa, O. 1978. Hyponatremia of hypothyroidism: appropriate suppression of antidiuretic hormone levels. *Arch. Intern. Med.* 138:820–22

84. Fanestil, D. D. 1977. Hypoosmolar syndromes. In *Disturbances in Body*

Fluid Osmolality, ed. T. E. Andreoli, J. J. Grantham, F. C. Rector Jr., 13:267–84. Bethesda, Md: Am. Physiol. Soc. 349 pp.

85. Kleeman, C. R., Rubini, M. E., Lamdin, E., Epstein, F. H. 1955. Studies on alcohol diuresis. II. The elevation of ethyl alcohol as an inhibitor of neurohypophysis. *J. Clin. Invest.* 34:448–55
86. Helderman, J. H., Vestal, R. E., Rowl, J. W., Tobin, J. D., Reubin, A., Robertson, G. L. 1978. The response of arginine vasopressin to intravenous ethanol and hypertonic saline in man: the impact of aging. *J. Gerontol.* 33:39–47
87. Tanay, A., Yust, I., Peresecenschi, G., Ambramov, A. L., Aviram, A. 1979. Long term treatment of the syndrome of inappropriate antidiuretic hormone secretion with phenyltoin. *Ann. Intern. Med.* 90:50–52
88. Miller, M. 1975. Inhibition of ADH release in the rat by narcotic antagonists. *Neuroendocrinology* 19:241–51
89. Kamoi, K., White, K., Robertson, G. L. 1979. Opiates elevate the osmotic threshold for vasopressin (VP) release in rats. *Clin. Res.* 27:254A
90. Miller, M., Moses, A. M. 1976. Drug-induced states of impaired water excretion. *Kidney Int.* 10:96–103
91. Forrest, J. N. Jr., Cohen, A. D., Torretti, J., Himmelhoch, J. M., Epstein, F. H. 1974. On the mechanism of lithium-induced diabetes insipidus in man and rat. *J. Clin. Invest.* 53:1115–23
92. White, M. G., Fetner, C. D. 1975. Treatment of the syndrome of inappropriate secretion of antidiuretic hormone with lithium carbonate. *N. Engl. J. Med.* 282:390–92
93. Forrest, J. N., Cox, M., Hong, C., Morrison, G., Bia, M., Singer, I. 1978. Superiority of demeclocycline over lithium in the treatment of chronic syndrome of inappropriate antidiuretic hormone. *N. Engl. J. Med.* 298:173–77
94. Tangedahl, T. N., Gan, G. T. 1972. Myocardial irritability associated with lithium carbonate therapy. *N. Engl. J. Med.* 287:867–69
95. Forrest, J. N. Jr. 1975. Lithium inhibition of cAMP-mediated hormones: a caution. *N. Engl. J. Med.* 292:423–24
96. Singer, I., Rotenberg, D. 1973. Demeclocycline induced nephrogenic diabetes insipidus: *in vitro* and *in vivo* studies. *Ann. Intern. Med.* 79:679–83
97. Wilson, D. M., Perry, H. O., Sams, W. M. Jr., Dousa, T. P. 1973. Selective inhibition of human distal tubular function by demeclocycline. *Curr. Ther. Res.* 15:734–40
98. Singer, I., Forest, J. N. Jr. 1976. Drug-induced states of nephrogenic diabetes insipidus. *Kidney Int.* 10:82–95
99. Cherril, D. A., Stote, R. M., Birge, J. R., Singer, I. 1975. Demeclocycline treatment in the syndrome of inappropriate antidiuretic hormone secretion. *Ann. Intern. Med.* 83:654–56

Ann. Rev. Med. 1980. 31:329–43
Copyright © 1980 by Annual Reviews Inc. All rights reserved

EFFECTIVENESS AND RISKS OF CONTRACEPTION

◆7364

W. Droegemueller, M.D. and R. Bressler, M.D.

University of Arizona, Health Sciences Center, Tucson, Arizona 85724

INTRODUCTION

The choice of contraceptive method made by an individual or a couple provides an interesting insight into their social, psychological, and sexual relationship. The physician serving as a family planning counselor should provide the individual or couple with a realistic appraisal of the benefits and risks of each method.

We define contraception to include any temporary or permanent method to prevent pregnancy resulting from a coital act. Regretfully, there have not been major innovations in contraceptive technology during the past few years. However, there have been interesting studies on the impact of current contraceptives, including the effects of sterilization operations on the health of both males and females.

The safe, simple, inexpensive, totally effective, easily reversible contraceptive without side effects and not used at the time of the sexual act—in short, the ideal contraceptive—is yet to be developed. In this report, we outline effectiveness and standard risks. Rather than an all-inclusive review, we highlight recent improvements in standard contraceptives and express our thoughts concerning current risk-benefit ratios.

STERILIZATION

Recently, sterilization has become the most popular method of birth control in the United States. The dramatic and almost geometric rise in the number of sterilization procedures for both the female and the male partner is related to a complex interplay of changes in our cultural attitudes. Equal rights for women, relaxation of restrictions and barriers, emphasis on seri-

0066-4219/80/0401-0329$01.00

ous medical complications associated with oral contraceptives and intrauterine devices, and advances in surgical techniques have liberalized the thinking and practices of patients and physicians alike.

Female Sterilization

Ten years ago, approximately 80% of surgical sterilizations were vasectomies. Recently, the surgical pendulum has swung and now over 60% of sterility operations are performed on the female. This trend is expected to continue for the next few years.

The benefits of tubal ligation are obvious, for few operations offer more relief from anxiety and have such a high chance of success. Failure rates with tubal ligation range from one in 200 to one in 1000 cases, depending on the type of operation performed.

Standard textbooks dwell on the immediate risks of tubal surgery—namely, anesthetic complications and electrosurgical damage to the intestines or surrounding tissues. Advances in laparoscopic technique, including the use of either silastic bands or spring-loaded clips, have virtually eliminated the inadvertent "burns" associated with electrocautery. Also, because clips and bands produce less tissue damage, it is anticipated that there will be a higher potential for reversibility.

In the past, discussions regarding long-term sequelae of tubal ligation centered on the occasional patient who regretted her decision and on failure of the original operation. However, recent literature has emphasized a specific syndrome that develops subsequent to tubal ligation. The major components of the "post-tubal syndrome" are dysfunctional uterine bleeding and pelvic pain. Most often the abnormal bleeding is menorrhagia and the pelvic pain includes dysmenorrhea and dyspareunia.

The true incidence of this syndrome is difficult to establish, but is estimated to be 10–33% in women who have had laparoscopic tubal fulguration. Many authors have yet to recognize the relationship between the previous tubal surgery and the subsequent pelvic symptoms. Studies now in progress will help to clarify the frequency of this long-term morbidity. It is our opinion that a definite syndrome does exist. A small percentage of previously asymptomatic women develop pelvic complaints in the months following tubal ligation. Often, these chronic symptoms are severe enough that the patient demands further surgery.

It is important to differentiate between this syndrome and that of a psychosomatic etiology. However, we have been impressed by the specificity with which patients who have had the "post-tubal syndrome" describe their symptoms. The pelvic pain that these women experience is usually related to movement or physical activity. Upon pelvic examination, the uterus is freely mobile and normal in size. Pelvic tenderness is most prominent in the parametrial areas. The menorrhagia is usually a reflection of ovulatory cycles, and the pain is not relieved by oral contraceptives or other medical management. Hysterectomy is successful in alleviating the symptoms.

Colleagues who remain skeptical of the association postulate that the syndrome is overdiagnosed and that it is merely a phenomenon of aging. Recent oral contraceptive use or an intervening pregnancy could cause a woman to forget her previous similar symptoms and result in a false assumption of cause and effect.

The etiology of the "post-tubal syndrome" remains as obscure as prospective studies substantiating its incidence. One theory suggests that the pain may be related to intermittent torsion of the fimbria and/or the ovary. Other theories postulate that the symptoms result from compromise of the delicate vascular network between the uterus and the ovaries. We believe that the pain is secondary to retrograde menstruation into the closed segments of the fallopian tubes. Observations made during laparoscopy performed at the time of the menstrual flow demonstrate that approximately 50% of all women experience retrograde menstruation. It is possible that this normal physiologic variant is the precipitating factor in the development of the pelvic pain.

If 15% of women develop this syndrome, 85% do not. Therefore, it seems illogical to use anticipation of the "post-tubal syndrome" as an overriding argument in proposing hysterectomy as the operation of choice for sterilization. However, patients are counselled about this possibility as part of preoperative informed consent. Adequate informed consent invariably requires more time than does the indicated sterilization procedure itself. The hallmarks of "informed consent" in regard to sterilization are that a mature adult has made a voluntary decision after considering the risks and alternatives. It is important that there has been an opportunity to ask and answer all questions and a reasonable period of time between the initial decision and the surgical procedure. Presently, the federal government, activist groups, and lawyers are all involved in formulating the specific details of the doctor-patient preoperative discussion regarding sterilization.

Concurrent with the increase in numbers of female sterilizations, there has been a dramatic rise in the number of women requesting reversal of their sterilization operation. Unexpected changes in an individual's personal life and publicity surrounding improvements in microsurgical techniques have increased these numbers. Several studies have determined the "profile" of women requesting reversal of sterilization. Currently, it is estimated that 1–2% of women regret their initial decision to have a tubal ligation and seek an operation to restore their fertility.

In a report by Victor Gomel of 100 consecutive patients, the major reasons for requesting sterilization reversal were: changes in marital status, 65%; crib death, 17%; and the desire for more children in the same marriage, 10% (21). Ironically, all 17 women who experienced the tragedy of crib death in their families had post-partum sterilization operations. Retrospectively, many patients stated that there were significant marital problems prior to and at the time of the operation. Hindsight is rarely cloudy; however, a commitment to sterilization made late in pregnancy or the puerperium is like "adopting religion on the battlefield." It is a commitment that may not last throughout the reproductive life of the individual. Permanent sterilization is a major decision that should be made at a time of psychological tranquility.

On-going research in female sterilization focuses on two major elements: the development of an easily reversible sterilization operation and transcervical methods of tubal occlusion. Two novel techniques of transcervical sterilization may result in a method with ease of reversal. Hosseinian et al have developed a uterotubal plug that is inserted via a hysteroscope (22). Pregnancies were obtained in baboons following removal of the device. Similarly, the silastic plug developed by Erb had encouraging results in animal experiments and is currently being tested in preliminary clinical trials (14). Several groups are investigating caps, hoods, socks, or plugs secured around the fimbria. In some experiments, the fimbria are placed in a protected environment extraperitoneal, between the leaves of the broad ligament. Results are preliminary but promising.

Other investigators have focused their interest on transcervical outpatient sterilization. These include uterine injections of sclerosing agents such as methylcyanoacrylate or quinacrine, and thermal damage to the endometrium and uterotubal junction by cryosurgery, electrocautery, or laser beam. Research continues on these innovative methods, but none has achieved acceptable rates of bilateral tubal obstruction. Reversibility is obviously a low priority in developing a new modality for outpatient sterilization.

Male Sterilization

Both scientifically and practically, vasectomy is simpler with less immediate morbidity, less expense, and easier reversibility than tubal ligation. The failure rate is slightly less with male sterilization. Most importantly, the success of vasectomy can be measured by a laboratory test. Because of the longer reproductive lifespan and anxiety of sexual performance, males experience a higher incidence of psychological sequelae following vasectomy than do women following tubal ligation.

As with female sterilization, the number of men requesting reversal of their previous sterilization procedure has increased recently. The urologists have been quicker than their gynecological peers to adapt microsurgical techniques for sterilization reversal. Therefore, the overall pregnancy rate with vasovasostomy is far superior to operations on the female. Pregnancy rates of 60–75% in unselected series are now common.

There have been two major areas of research with regard to vasectomy in the past five years. The efforts of several groups and millions of dollars have not successfully developed a reversible intravasal device. The fantasy was appealing but it has not become a reality. Recent interest has focused on the possibility of producing vas occlusion without a surgical incision. Injections of sclerosing solutions have produced asospermia in individual patients (20).

Most alarming is a report that vasectomy may be a surgical model for autoimmune disease. Nancy Alexander (1) postulated that the immunological response to sperm antigens which often accompanies vasectomy may accelerate the development of atherosclerotic cardiovascular changes. She demonstrated that diet-induced atherosclerosis developed more extensively in vasectomized cynomolgus monkeys than in sham-operated control monkeys fed the same diet. Complement and immunoglobulins were associated with the atherosclerotic plaques in some of the monkeys. (Important references on sterilization include 3, 10, 12, 13, 23, 24, 33, 36–38, 42, 43, 53, 59, 60.)

STEROIDAL CONTRACEPTIVES

No other development in medical history has had such a revolutionary impact on society as the introduction of oral contraceptives. For the past 20 years the steroidal oral contraceptives have proven to be among the most effective pharmacologic products ever marketed. The theoretical effectiveness is one pregnancy per 1000 woman years of use, and the actual use effectiveness is 5–8 pregnancies per 1000 woman years of use. They are immensely popular because of this effectiveness and because their use is removed in time from the sexual act. Currently, it is estimated that 10 million women in the United States and 60 million women in the world are using oral contraceptives. Their popularity continues to increase in the face of widespread publicity concerning the serious complications, and women having to endure the minor side effects caused by the "pseudopregnancy."

Because so many women use them, oral contraceptives are among the most thoroughly studied drugs. Many review articles document the various adverse effects that have been recorded in the literature. The mortality risk from the pill in women under the age of 35 is three in 100,000. If one considers risks of life, the chance of the same women being killed in an automobile accident or murdered on the streets of a metropolitan city is approximately 20 times greater.

Few authors comment on the beneficial effects of oral contraceptives. Although the list of positive side effects does not compare in length or breadth with the harmful aspects, it is worth discussing. Women taking oral contraceptives have fewer menstrual disorders, including a relief from dysmenorrhea, and fewer hospital admissions for diagnostic dilatation and curettage for abnormal bleeding. Oral contraceptives inhibit the development of benign breast disease and decrease the incidence of functional ovarian cysts. Lastly, a statistical decrease has been reported in some medical illnesses including peptic ulcers and rheumatoid arthritis.

Recently, there has been an emphasis on reformulation with lower dosages in an effort to minimize the side effects and complications of oral contraceptives. It has been suggested that oral contraceptives should contain at least 30 μg of ethinyl estradiol. Lower dosages of estrogen will result in either an excessive incidence of abnormal bleeding or a higher pregnancy rate.

Obviously, a physician takes into account the absolute contraindications to oral contraceptives, such as previous history of thromboembolic disease, deep-vein thrombophlebitis and cerebrovascular disease, active liver disease or impaired liver function, estrogen-dependent neoplasia, and suspected or known pregnancy. Similarly, relative contraindications should be considered, such as migraine headaches, oligomenorrhea, hypertension, 40 or more years of age, epilepsy, sickle cell disease, and history of idiopathic jaundice of pregnancy. However, the decision to use the pill is very complex and often involves not only medical but social, moral, and economic issues. Therefore each patient must be considered individually.

Rather than reiterating previous reports of the side effects of oral contraceptives, we discuss here changes in clinical practice that may reduce adverse effects. It is important to include a family history of hypercoagulability as a relative contraindication to oral contraceptives, because the pill causes all women to be hypercoagulable. However, some of the intravascular coagulation problems can be prevented if the patient is encouraged to discontinue oral contraceptives prior to elective surgery, cardiac catherization, and other procedures that predispose to thrombophlebitis. Another relative contraindication not stressed in either the gynecological or medical literature is depression. Obviously, it is difficult to precisely measure the depressive affect in any person's life. However, the pill can definitely increase depression, and many serious psychiatric sequelae have followed the initiation of oral contraceptives.

Patients should be instructed that it is important to discontinue oral contraceptives about three months prior to conceiving their next pregnancy. There are soft data that chromosomal abnormalities are higher in pregnancies that begin the month following discontinuence of oral contraceptive therapy. Of more clinical significance is the woman and her physician both being certain of the time of conception. In modern obstetrics there are only two therapeutic modalities; bed rest, which improves uterine blood flow, and termination of the pregnancy prior to the expected date of confinement. A pregnancy with unknown dates complicated by intrauterine growth retardation, for example, becomes a laboratory dystocia requiring multiple ultrasounds, estriols, amniotic fluid analysis, and stress testing of the fetal heart rate.

Most physicians instruct their patients to return three months from the first visit after instituting oral contraceptives. This allows the physician to discover idiosyncratic reactions such as hypertension or growth of myomas. Because of the association of multiple congenital anomalies, the VACTERL syndrome associated with the ingestion of synthetic progestins during early pregnancy, it is prudent to avoid these compounds not only in diagnostic tests for amenorrhea but also in the treatment of dysfunctional uterine bleeding. Both of the latter admonitions are included for patients who are "sexually active" and possibly could be pregnant. Progesterone in oil intramuscularly is an acceptable alternative as part of the endocrine investigation of a woman with amenorrhea.

Liver adenomas are the most prominent complication of oral contraceptives that have been related to chronicity of use. The incidence of most other risks including the "post-pill amenorrhea" are not determined by how long a woman takes oral contraceptives. Regretfully, this clinical fact is not appreciated by all practitioners. Therefore, many women have been instructed to discontinue their pills for two to three months every couple of years, the "periodic rest period." This unwarranted clinical practice may help to alleviate iatrogenic guilt or a small obstetrical practice, but is not based on sound clinical studies.

A recent publication documented an increased conversion rate from dysplasia to carcinoma in situ of the cervix in women on the pill as compared to women using the intrauterine device (48). The probability was six-fold greater that those on the pill would have carcinoma in situ within the seven-year time limit of the study. This finding must be put in perspective of the entire risk-benefit ratio. At worst, cervical cancer is a skin cancer with corresponding excellent cure rates. Denial of the pill to a woman just because she had dysplasia of the cervix is analogous to banning the bikini for fear of skin cancer in the sunbelt.

With the gradual reduction of estrogen content in oral contraceptives, some women have reached the minimum dosage necessary for adequate contraception. Recently, there have been several different drugs implicated in drug interactions with the pill. Rifampicin increases the rate of hepatic metabolism, thereby reducing plasma concentrations of estrogen. Consequently, breakthrough bleeding and pregnancies have been reported. There has also been an occasional pregnancy associated with low-dose oral contraceptives and chloramphenicol, phenytoin, and other antiepileptic drugs.

The progestin-only mini-pill has not gained much popularity in the United States. The pregnancy rate of 2–5 per 100 woman years and associated disruption of the normal bleeding cycle, which is replaced by amenorrhea and irregular episodes of spotting and bleeding, produce too much anxiety for the average female. It should be used primarily during lactation and in patients who have complications related to the estrogen component of the pill such as growth of uterine myomas.

Depomedroxyprogesterone given in injections of 150 mg every three months is a popular contraceptive outside of this country. Recent evidence of endometrial carcinoma in two monkeys receiving this drug will probably further depress any hopes of FDA approval.

Research has centered on innovative ways of slowly releasing steroids. Silastic vaginal rings are being investigated that would either release both estrogen and progestin or progestin alone. The rings would eliminate the necessity of the woman remembering to take her pill each day. However, similar to the mini-pill, the menstrual pattern would change. Currently, there are several clinical trials of injecting or implanting steroids in biodegradable rods or pellets for prolonged contraceptive action. (Important references on oral contraceptives include 5, 6, 8, 11, 18, 25, 27, 30, 35, 39, 46, 47 54–58.)

INTRAUTERINE DEVICES

The intrauterine device (IUD) as a contraceptive had a long history of medical nonacceptance in the United States before 1962. The IUD fell into clinical disfavor in the 1930s because of the concern over serious pelvic infections associated with its use. With the development of new polyethylene devices of high tissue tolerance, most of these objections disappeared. A World Health Organization report in 1968 stated that pelvic inflammatory disease associated with the IUD was not a significant problem and, at that time, there was no evidence that the incidence of pelvic infection was greater in IUD wearers than in the general population. Ten years ago, the IUD was gaining in popularity at a rate greater than any other contraceptive. However, because of recent reports concerning increasing incidence of pelvic infection, especially serious sequelae such as unilateral tubo-ovarian abscess and pelvic actinomycoses, history is repeating itself.

The IUD has several benefits. The failure or pregnancy rate is 2–3 per 100 woman years. Unusual for most contraceptives, the theoretical effectiveness and use effectiveness are virtually the same. Continuous motivation or performance is not necessary to successfully use the method. Also, the effect is local and there are no associated systemic metabolic effects. In considering mortality rates related specifically to the contraceptive method, oral contraceptives are three times more dangerous than the IUD. Conversely, if one considers hospitalizations directly related to the contraceptive method, the incidence associated with the IUD is 5–10 times higher than the pill.

Because of infectious complications and their relationship to future fertility, many clinics do not like to insert an IUD into nulliparous women. However, the IUD is well suited for older multiparous women who have completed their families but are not prepared to have a surgical sterilization. Older parous women tolerate the device with fewer expulsions and a higher continuation rate.

The mechanism of action of the plastic IUD is related to the sterile inflammation produced by the foreign body in the uterus. The larger the surface area of the device, the greater the inflammatory response and the lower the failure rate. The latest generation of IUDs are smaller and medicated with either copper or progesterone. The advantages of medicated IUDs are that, because of their smaller size, they are more easily tolerated and produce fewer local symptoms, especially in women with smaller uteri. The copper gradually diffuses from the device and, therefore, a new IUD should be inserted after three years. Progesterone is released at a rate of 65 μg daily to the endometrium and enhances the local contraceptive effect without producing levels of hormone that cause systemic effects. The progesterone IUD should be replaced after 12 months of use.

Uterine perforation is the foremost immediate complication of the IUD. The exact incidence of this complication is unknown since many perforations are not discovered, but it is quoted as being between one in 400 and one in 1000 insertions. The patient whose IUD appendage or string is lost must be investigated for this complication. Ultrasound visualization or both anterior-posterior and lateral X rays with a metal sound in the uterus will establish the diagnosis. It is imperative to remove an intraperitoneal copper device because the peritoneum reacts adhesively to the copper.

Pregnancy is an alarming adverse effect to the majority of women who experience this unanticipated complication. The IUD is definitely not teratogenic. Because of its extraamniotic location, the IUD can facilitate intrauterine infection and result in premature rupture of the membranes. A physician should remove the device by pulling on the vaginal appendage as soon as pregnancy is suspected. This apparently illogical maneuver decreased the fetal loss from 48% when the device was left in situ to a 20–30% abortion rate following removal of the IUD.

Three nuisance side effects may cause patients to select another contraceptive method; expulsion, bleeding, and pelvic pain. Expulsion occurs in 5–15% of women and depends directly on the size of the device inserted. Bleeding is believed to be related to a local increase in the fibrinolytic system of the endometrium. Abnormal bleeding with the IUD has been treated successfully with aminocaproic acid and tranexamic acid and other similar synthetic antifibrinolytic agents. Because of the systemic gastrointestinal toxicity of these drugs, attempts are now being made to incorporate these agents into the medicated IUDs.

The pelvic pain is usually related to the release of prostaglandins by the endometrial stroma. Oral prostaglandin synthetase inhibitors, aspirin and nonsteroidal anti-inflammatory agents, effectively relieve the dysmenorrhea associated with an intrauterine foreign body. No investigator has been bold enough to treat women with prostaglandin synthetase inhibitors during the entire cycle, fearing that the contraceptive action of the sterile inflammation might be disturbed. During normal menstruation a woman without an IUD loses approximately 38 ml of blood. The average woman with an IUD will lose 80 ml per period, and an occasional patient will bleed considerably more. Therefore, the IUD should be considered in the differential diagnosis of a woman with an iron deficiency anemia.

Recent physician disenchantment with the IUD centers on the infectious complications. The incidence of acute pelvic inflammatory disease is increased 5- to 10-fold by the presence of an IUD regardless of the socioeconomic status of the patients. Major pelvic infections resulting in death have been reported in both pregnant and nonpregnant women.

Not only is there an increase in common bilateral salpingitis, but unusual and rare pelvic infections are now being seen with alarming frequency in IUD users. In 1975, a report of unilateral tubo-ovarian abscesses associated with the IUD was published (52). The women had a syndrome characterized by 2–5 weeks of prodromal symptoms including vague lower abdominal pain, pelvic tenderness, and dyspareunia. If untreated this insidious infection can develop into eventual loss of the tube and ovary. It is believed that the IUD should be removed when prodromal symptoms arise. In the absence of a definite pelvic mass the patient may be treated with antibiotics as an outpatient. However if a definite pelvic mass is palpated, the patient should be admitted to the hospital and receive intravenous antibiotic therapy. Early surgical therapy should be considered if the mass does not respond to medical management.

Removal of an asymptomatic IUD may precipitate an inflammatory response and secondary infection during concurrent elective pelvic surgery such as tubal ligation.

Pelvic actinomycoses, another rare but potentially serious pelvic infection associated with the IUD, is being reported in increasing numbers. This organism is not part of the normal vaginal flora although it is considered normal flora in the oropharynx and bowel. In a study at Johns Hopkins, 25% of symptomatic IUD users had actinomycetes organisms in Fast, Papanicolaou smears (31, 45). In a small series, 80% of asymptomatic women with IUDs and actinomycetes identified on Pap smears subsequently developed symptoms, including vaginitis, abnormal bleeding, and actue pelvic inflammatory disease. The proper management of women with actinomycetes on smears is unclear, but some gynecologists are removing the device and treating the patient with penicillin.

The basic physiologic action of the intrauterine device is ineffective in protecting against ectopic pregnancy. There is a 12-fold increase in risk of ectopic pregnancy between IUD users and women using mechanical contraception. This incidence continues to increase with chronicity of IUD usage. This tendency may possibly continue to a lesser extent even after the IUD is removed. One of eight ectopic pregnancies associated with the IUD is a primary ovarian pregnancy. The most tragic aspect of this relationship is that the symptoms normally produced by an IUD and the constellation of classic symptoms associated with an ectopic pregnancy are so similar that the diagnosis of the ectopic is often delayed. Lower abdominal and pelvic pain, spotting, and signs of pelvic inflammatory disease are often attributed to the IUD and the diagnosis of ectopic not even entertained. In one series, 50% of patients had their IUDs removed prior to even considering the diagnosis of ectopic pregnancy. Obviously, this delay in awareness and diagnosis can be life threatening.

Two quotes from eminent gynecologists will help to underline current thinking regarding the IUD. Dr. E. Stewart Taylor wrote, "I think the IUD should not be used by women who have not had their family unless there is some compelling social indication" (52). Dr. Robert Barter stated, "Categorically and unequivocally . . . in a patient who has an IUD or who may have an IUD and who has pelvic pain with or without fever the IUD must be removed as the first step in any therapy, and any therapy without its removal is meaningless, of no value, and worthless. It is amazing that this concept still does not have universal acceptance" (4). (Important references on the IUD include 2, 9, 15–17, 26, 32, 34, 40, 41, 44, 49–51.)

BARRIER CONTRACEPTIVES

Successful use of barrier contraceptives depends on a highly motivated individual or couple, for effectiveness depends on proper placement of the barrier immediately prior to the sexual act. Therefore, in studies of all forms of vaginal contraception and condoms, there is a wide discrepancy between theoretical and use effectiveness.

There has not been a medical change in the condom since the days of Casanova. The theoretical effectiveness is 3 per 100 woman years while the use effectiveness is variously quoted between 12 and 20 per 100 woman years. The only side effect is a rare allergic reaction to rubber. This can be alleviated by having the man switch to a condom made of collagen or the skin condom, made from the caecum of lambs. Many women believe that the male partner should assume the responsibility for contraception. Manufacturers have responded with "new improved" condoms that are thinner, lubricated, ribbed, and multicolored. The impact of these changes has not been chronicled in the medical literature

The vaginal diaphragm has a theoretic effectiveness of 2 per 100 woman years, and a closely monitored study in highly motivated individuals resulted in a similar use effectiveness. However, large-scale investigations have documented pregnancy rates of 12–18 per 100 woman years. Rarely, a woman may be allergic to latex or the companion spermicidal cream or jelly. Because of the adverse publicity surrounding oral contraceptives and the intrauterine device, private industry has been innovative in new methods of delivering spermicides to the vagina. Presently, there are many new creams, jellies, suppositories, foaming tablets, and removable and biodegradable sponges that deliver Nonoxynol-9 to the upper vagina. The advantages in the market place are that most of the products are smaller and have a more pleasant odor than a diaphragm. The immediate side effects include a warm sensation in the vagina, as some of these products effervesce, and a "gritty" feeling if there is not adequate vaginal lubrication. Since virtually all of these new barrier methods contain the same spermicide, it is difficult to visualize significant changes in pregnancy rates. Some scientists are beginning to question the toxicity of the detergents used to lyse spermatozoa. Their ability to disintegrate cells or subcellular membrane structures such as the lysosomal vacuole or endoplasmic reticulum is well established. Because spermicidal agents in the upper vagina are rapidly absorbed, there is the possibility of a systemic effect. Preliminary toxicologic studies in rats and rabbits indicate liver and renal damage. These results dampen enthusiasm for chronically releasing Nonoxynol-9 from a vaginal ring or similar device (7, 19, 28, 29).

SUMMARY

Reproductive biology has made significant advances as demonstrated by the recent birth of an egg fertilized in vitro. But even in the era of "test-tube babies," prevention of pregnancy is not without associated risks. The ideal contraceptive that is safe, efficacious, and reversible is still not a reality. The lowest mortality rates are associated with barrier contraceptives with a back-up of early abortion when the method fails. The relative risks, efficacy, side effects, and costs of various available contraceptives are summarized in Table 1.

Table 1 Risk-benefit ratio of contraceptives

Method	Failure rate or effectiveness	Major side effects	Minor side effects	Approximate cost
Female sterilization	1 in 200 to 1 in 1000	post-tubal syndrome	none	$850
Male sterilization	1 in 1000	possible atherosclerosis	none	$175
Oral contraceptives	0.5 per 100 woman years	thrombophlebitis pulmonary embolus cerebral thrombosis	pseudo-pregnancy	$ 75/yr
IUD	2–3 per 100 woman years	uterine perforation pelvic infections	uterine bleeding and cramping	$ 75/yr
Barrier methods	3–20 per 100 woman years	none	allergic reaction nuisance to use	$ 30/yr

The choice of a contraceptive method is a very personal one and involves many facets of daily life. This choice involves various risk-benefit ratios, though the eventual decision often is not a scientific one. However, it is the physician's responsibility to provide the patient with the factual background for his or her contraceptive choice.

Literature Cited

1. Alexander, N. J. 1978. Vasectomy increases the severity of diet-induced atherosclerosis in *Macaca fascicularis. Science* 201:538–41
2. Anderson, A. B. M., Haynes, P. J., Guillebaud, J., Turnbull, A. C. 1976. Reduction of menstrual blood-loss by prostaglandin-synthetase inhibitors. *Lancet* 1:774–76
3. Baggish, M. S., Lee, W. K., Miro, S. J., Dacko, L., Cohen, G. 1979. Complications of laparoscopic sterilization. Comparison of 2 methods. *Obstet. Gynecol.* 54:54–59
4. Barter, R. H. 1975. Discussion of paper by Taylor, E. S. *Am. J. Obstet. Gynecol.* 123:342
5. Chez, R. A. 1978. Proceedings of the symposium "Progesterone, Progestins, and Fetal Development." *Fertil. Steril.* 30:16–26
6. Christopherson, W. M., Mays, E. T., Barrows, G. 1977. A clinicopathologic study of steroid-related liver tumors. *Am. J. Surg. Pathol.* 1:31–41
7. Chvapil, M., Eskelson, C. D., Droegemueller, W., Ulreich, J. B., Owen, J. A., Ludwig, J. C., Stiffel, V. W. 1979. New data on the pharmacokinetics of Nonoxynol-9. In *New Developments in Vaginal Contraception,* ed. G. I. Zatuchni, A. J. Sobrero, J. J. Speidel, J. J. Sciarra. Hagerstown, Md: Harper & Row. In press
8. Coskey, R. J. 1977. Eruptions due to oral contraceptives. *Arch. Dermatol.* 113:333–34
9. Dische, F. E., Burt, J. L., Davidson, N. J. H., Puntambekan, S. 1974. Tubo-ovarian actinomycosis associated with intrauterine contraceptive devices. *J. Obstet. Gynaecol. Br. Commonw.* 81:724–29
10. Droegemueller, W., Greer, B. E., David, J. R., Makowski, E. L., Chvapil, M., Pollard, A. 1978. Cryocoagulation of the endometrium at the uterine cornua. *Am. J. Obstet. Gynecol.* 131:1–7
11. Durand, J. L., Bressler, R. 1979. Clinical pharmacology of the steroidal oral contraceptives. *Adv. Intern. Med.* 24: 97–126
12. Emens, J. M., Olive, J. E. 1978. Timing of female sterilization. *Br. Med. J.* 2:1126

13. Engel, T. 1978. Laparoscopic sterilization: Electro-surgery or clip application? *J. Reprod. Med.* 21:107–10
14. Erb, R. A. 1976. Silastic: A retrievable custom-molded oviductal plug. In *Advances in Female Sterilization Techniques,* ed. J. J. Sciarra, W. Droegemueller, J. J. Speidel, pp. 259–72. Hagerstown, Md: Harper & Row
15. Eschenbach, D. A. 1976. Acute pelvic inflammatory disease: etiology, risk factors and pathogenesis. *Clin. Obstet. Gynecol.* 19:147–69
16. Everett, E. D., Reller, L. B., Droegemueller, W., Greer, B. 1976. Absence of bacteremia after insertion or removal of intrauterine devices. *Obstet. Gynecol.* 47:207–9
17. Faulkner, W. L., Ory, H. W. 1976. Intrauterine devices and acute pelvic inflammatory disease. *J. Am. Med. Assoc.* 235:1851–53
18. Fisch, I. R., Frank, J. 1977. Oral contraceptives and blood pressure. *J. Am. Med. Assoc.* 237:2499–2503
19. Free, M. J., Alexander, N. D. 1976. Male contraception without prescription: a reevaluation of the condom and coitus interruptus. *Public Health Rep.* 91:437–45
20. Freeman, C. 1975. Preliminary human trial of a new male sterilization procedure: vas sclerosing. *Fertil. Steril.* 26:162–66
21. Gomel, V. 1978. Profile of women requesting reversal of sterilization. *Fertil. Steril.* 30:39–41
22. Hosseinian, A. H., Lucero, S., Kim, M. H. 1976. Hysteroscopic implantation of uterotubal junction blocking devices. See Ref. 14, pp. 169–79
23. Hulka, J. F., Mercer, J. P., Fishburne, J. I., Kumarasamy, T., Omran, K. F., Phillips, J. M., Lefler, H. T., Lieberman, B., Lean, T. H., Pai, D. N., Koetsawang, S., Castro, V. W. 1976. Spring clip sterilization: One-year follow-up of 1,079 cases. *Am. J. Obstet. Gynecol.* 125:1039–43
24. Hulka, J. F., 1977. Current status of elective sterilization in the United States. *Fertil. Steril.* 28:515–20
25. Jick, H., Porter, J., Rothman, K. J. 1978. Oral contraceptives and nonfatal stroke in healthy young women. *Ann. Intern. Med.* 89:58–59
26. Kahn, H. S., Tyler, C. W. 1975. IUD-related hospitalizations United States and Puerto Rico–1973. *J. Am. Med. Assoc.* 234:53–56
27. Kaplan, N. M. 1978. Cardiovascular complications of oral contraceptives. *Ann. Rev. Med.* 29:31–40
28. Lane, M. E., Arceo, R., Sobrero, A. J. 1976. Successful use of the diaphragm and jelly by a young population: Report of a clinical study. *Fam. Plann. Perspect.* 8:81–86
29. Lane, M. E. 1977. Benefits and risks of diaphragm and condom use. In *Risks, Benefits, and Controversies of Fertility Control,* ed. J. J. Sciarra, G. I. Zatuchni, J. J. Speidel, pp. 547–55. Hagerstown, Md: Harper & Row
30. LiVolsi, V. A., Stadel, B. V., Kelsey, J. L., Holford, T. R., White, C. 1978. Fibrocystic breast disease in oral-contraceptive users. *N. Engl. J. Med.* 299:381–85
31. Luff, R. D., Gupta, P. K., Spence, M. R., Frost, J. K. 1978. Pelvic actinomycosis and the intrauterine contraceptive device. A cyto-histomorphologic study. *Am. J. Clin. Pathol.* 69:581–86
32. Mead, P. B., Beecham, J. B., Maeck, J. V. S. 1976. Incidence of infections associated with the intrauterine contraceptive device in an isolated community. *Am. J. Obstet. Gynecol.* 125:79–82
33. Neil, J. R., Hammond, G. T., Noble, A. P., Rushton, L., Letchworth, A. T. 1975. Late complications of sterilization by laparoscopy and tubal ligation. *Lancet* 2:699–700
34. Niebyl, J. R., Parmley, T. H., Spence, M. R., Woodruff, J. D. 1978. Unilateral ovarian abscess associated with the intrauterine device. *Obstet. Gynecol.* 52:165–68
35. Nora, J. J., Nora, A. H., Blu, J., Ingram, J., Fountain, A., Peterson, M., Lortscher, R. H., Kimberling, W. J. 1978. Exogenous progestogen and estrogen implicated in birth defects. *J. Am. Med. Assoc.* 240:837–43
36. Pauerstein, C. J. 1978. From fallopius to fantasy. *Fertil. Steril.* 30:133–40
37. Richart, R. M., Neuwirth. R. S., Bolduc, L. R. 1977. Single-application fertility-regulating device: Description of a new instrument. *Am. J. Obstet. Gynecol.* 127:86–90
38. Ringrose, C. A. D. 1974. Post tubal ligation menorrhagia and pelvic pain. *Int. J. Fertil.* 19:168–70
39. Robertson, Y. R., Johnson, E. S. 1976. Interactions between oral contraceptives and other drugs. A review. *Curr. Med. Res. Opin.* 3:647–61
40. Rosenfield, A. 1978. Oral and intrauterine contraception: A 1978 risk assess-

ment. *Am. J. Obstet. Gynecol.* 132:92–106

41. Rybo, G. 1978. The IUD and endometrial bleeding. *J. Reprod. Med.* 20:175–82
42. Schneider, J. 1973. Counselling for sterilization. *Obstet. Gynecol.* 42:788–84
43. Sciarra, J. J., Zatuchni, G. I., Speidel, J. J., eds. 1977. *Reversal of Sterilization.* Hagerstown, Md: Harper & Row. 294 pp.
44. Smith, M. R., Soderstrom, R. 1976. Salpingitis: a frequent response to intrauterine contraception. *J. Reprod. Med.* 16:159–62
45. Spence, M. R., Gupta, P. K., Frost, J. K., King, T. M. 1978. Cytologic detection and clinical significance of *Actinomyces israelii* in women using intrauterine contraceptive devices. *Am. J. Obstet. Gynecol.* 131:295–98
46. Spencer, J. D., Millis, R. R., Hayward, J. L. 1978. Contraceptive steroids and breast cancer. *Br. Med. J.* 1:1024–26
47. Steel, J. M., Duncan, L. J. P. 1978. Serious complications of oral contraception in insulin-dependent diabetics. *Contraception* 17:291–95
48. Stern, E., Forsythe, A. B., Youkeles, L., Coffelt, C. F. 1977. Steroid contraceptive use and cervical dysplasia: Increased risk of progression. *Science* 196:1460–62
49. Targum, S. D., Wright, N. H. 1974. Association of the intrauterine device and pelvic inflammatory disease: a retrospective pilot study. *Am. J. Epidemiol.* 100:262–71
50. Tatum, H. J., Schmidt, F. H. 1977. Contraceptive and sterilization practices and extrauterine pregnancy: A realistic perspective. *Fertil. Steril.* 28:407–21
51. Tauber, P. F., Wolf, A. S., Herting, W., Zaneveld, L. J. D. 1977. Hemorrhage induced by intrauterine devices: control by local proteinase inhibition. *Fertil. Steril.* 28:1375–77
52. Taylor, E. S., McMillan, J. H., Greer, B. E., Droegemueller, W., Thompson, H. E. 1975. The intrauterine device and tuboovarian abscess. *Am. J. Obstet. Gynecol.* 123:338–42
53. Thomson, P., Templeton, A. 1978. Characteristics of patients requesting reversal of sterilization. *Br. J. Obstet. Gynaecol.* 85:161–64
54. Vessey, M. P., Doll, R. 1976. Evaluation of existing techniques. Is "the pill" safe enough to continue using? *Proc. R. Soc. Lond. Ser. B.* 195:69–80
55. Vessey, M. P., Wright, N. H., McPherson, K., Wiggins, P. 1978. Fertility after stopping different methods of contraception. *Br. Med. J.* 1:265–67
56. Vessey, M. P., McPherson, K., Johnson, B. 1977. Mortality among women participating in the oxford/family planning association contraceptive study. *Lancet* 2:731–33
57. von Kaulla, E., Droegemueller, W., Aoki, N., von Kaulla, K. N. 1971. Antithrombin III depression and thrombin generation acceleration in women taking oral contraceptives. *Am. J. Obstet. Gynecol.* 109:868–73
58. Wood, J. R. 1977. Ocular complications of oral contraceptives. *Ophthalmic Semin.* 2:371–402
59. Yoon, I. B., King, T. M., Parmley, T. H. 1977. A two year experience with the falope ring sterilization procedure. *Am. J. Obstet. Gynecol.* 127:109–12
60. Zipper, J., Stacchetti, E., Medel, M. 1975. Transvaginal chemical sterilization: Clinical use of quinacrine plus potentiating adjuvants. *Contraception* 12:11–21

Ann. Rev. Med. 1980. 31:345–60
Copyright © 1980 by Annual Reviews Inc. All rights reserved

PERICARDITIS AND RENAL FAILURE

◈7365

Roger Renfrew, M.D., Theodore J. Buselmeier, M.D., and Carl M. Kjellstrand, M.D.

Division of Nephrology, Department of Medicine, University of Minnesota
Medical School, Minneapolis, Minnesota 55455

INTRODUCTION

The spectrum of pericardial disease in association with uremia has broadened with the therapeutic advances for end-stage renal failure. The pericarditis that Bright initially described in 1836 (1) is most certainly a different entity etiologically, pathologically, and therapeutically from the pericarditis that occurs in well-dialyzed patients of today.

Pericarditis associated with end-stage renal disease can be categorized into uremic, dialysis, and constrictive pericarditis (Table 1) (2).

Uremic pericarditis occurs in patients with severe end-stage azotemia and uremia who have not previously received dialysis therapy. If the uremia is not treated, this form of pericarditis indicates a very poor prognosis. If the underlying uremia is corrected, the pericarditis will also respond to therapy.

Dialysis pericarditis occurs in two groups of patients who have already been stabilized on chronic hemodialysis. The first are patients who have a subclinical effusion detected by echocardiography, and second are patients who have clinically detectable pericarditis or pericardial effusion. These latter patients may or may not have a preceding exacerbating factor.

Constrictive pericarditis usually has a subacute or chronic progress, and therapeutic implications differ from the other two forms of pericardial disease.

Many authors describing the incidence, symptoms, results, and complications of therapy have not clearly distinguished between the groups of pericarditis as classified in Table 1. This probably accounts for wide variation in incidence or frequency figures. If one bases the incidence of pericarditis on routine screening of asymptomatic patients with echocardiography, the result is higher than when the incidence is based on severe clinical symp-

0066-4219/80/0401-0345$01.00

Table 1 Classification of pericarditis in uremic and dialyzed patients

- Uremic pericarditis
- Dialysis pericarditis
  - Subclinical
  - Clinical
    - With preceding stress
    - Without preceding stress
- Constrictive pericarditis

toms. Similarly, if success of therapy is based on treating asymptomatic patients with only echocardiography findings, the result of therapy will appear much better than if one includes only patients with severe symptomatic pericardial disease.

INCIDENCE

Uremic Pericarditis

In their review article, Comty et al (3) concluded, and others have reported (4–10), that the incidence of uremic pericarditis is decreasing. This is demonstrated both in inter- and intrainstitutional studies (11). Prior to the use of hemodialysis and in early dialysis series, uremic pericarditis was reported in 30–50% of patients (3–5). More recent series report an incidence of 3–19% (6, 8), which probably reflects earlier initiation of dialysis therapy. Thus, most of the patients who develop uremic pericarditis are patients who begin medical therapy late in the course of their uremia (8).

Dialysis Pericarditis

Pericarditis occurs in from 2–19% of patients who have been stabilized on chronic dialysis (3, 4, 6, 7, 9, 11–14). The incidence is partially a function of the diagnostic tools utilized. For example, autopsies on asymptomatic patients who died of other causes will add a number of cases to the clinical series of pericarditis (6), and echocardiography may diagnose an effusion in 11–28% of asymptomatic patients on chronic hemodialysis (15–18).

Dialysis pericarditis can develop at any time after a patient becomes stabilized on dialysis therapy (up to 50 months later), but 40% develop within the first three months (4, 6, 7, 13, 19).

CLINICAL PRESENTATION

Uremic Pericarditis

The symptoms of uremic pericarditis are less frequent and less severe than in dialysis pericarditis, and tamponade is rare (2). It is not, however, an occult problem—chest pain is reported in 32–83% of the patients (5, 20), rubs are heard in 57–100% of patients (4, 8, 10, 20), and cardiac tamponade occurs 7–13% of the time (8, 21–23).

Dialysis Pericarditis

It is difficult to obtain accurate data regarding the clinical features of dialysis pericarditis alone because they are usually combined with uremic pericarditis. The following data are based only on those series that consist of at least 60% dialysis pericarditis. The average and ranges of the reported series are found in the text and in the figures referred to in the text.

SYMPTOMS Approximately 30% of these pericarditis patients will have no symptoms (9, 24), although nonspecific symptoms such as malaise and weight loss (9) are common. A change in mental status also may be noted (3) (see Figure 1).

More specific complaints such as pain, which occurs in 75% of these patients, may implicate the cardiopulmonary system (6, 7, 9, 12, 24–26). The incidence of pain falls to 47–86% when series based on patients with intractable infusions or tamponade are excluded from evaluation (25, 26). Definite pleuritis or positional pain is reported in one half of these patients (7), and narcotics are required by one third of them (6, 7).

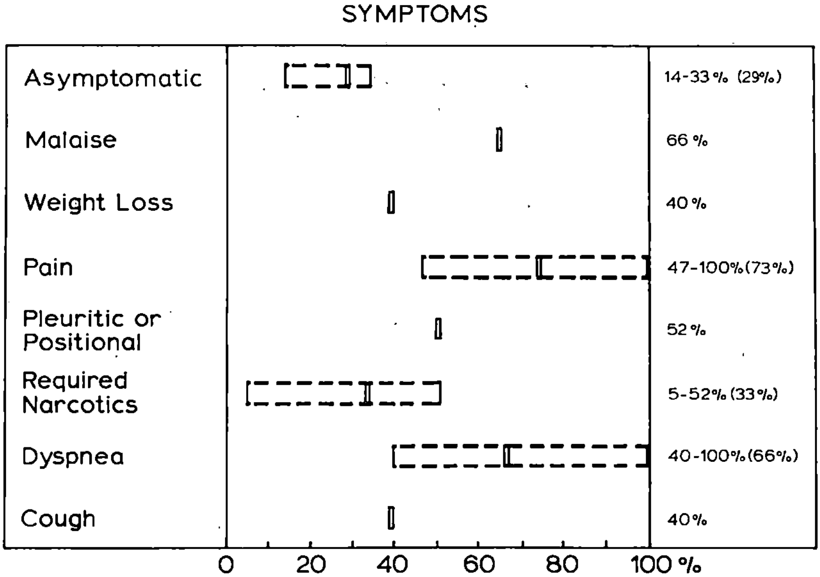


Figure 1 Symptoms reported by dialysis patients who have pericarditis. The most common symptoms are pain, dyspnea, and general malaise.

Cough and dyspnea are reported in at least 40% of the pericarditis patients (9). Dyspnea is found in 93% of those with tamponade or intractable effusion (25, 26).

PHYSICAL FINDINGS A low grade fever is found in 88% of the patients with dialysis pericarditis (6, 7, 9, 12, 25, 26), and in 96% of them pericardial friction rub can be heard (6, 9, 12, 24, 26, 27). Because the friction rub was a major criterion for the diagnosis of pericarditis, prior to the use of echocardiography, these figures probably overrepresent the frequency of pericardial friction rub with documentable pericardial disease. A pericardial friction rub, however, is highly specific for pericarditis (see Figure 2).

Congestive heart failure is present in about two thirds of the patients (9, 26). More specifically, a cardiac gallop is reported in 66% (9), and an increased jugular venous pressure is reported in 86% (27) of the patients.

LABORATORY Although leukocytosis is reported in 62% of the patients (6, 9, 12), it is usually only mild (6) (Figure 3). Roentgenographic findings of the chest are almost uniformly abnormal—cardiomegaly is found 99% of the time (9, 12, 24, 26, 27), although it is common in many cases of chronic renal failure (17). When a change in cardiac silhouette was looked for, it was found in 96% of the patients (6), and is obviously highly specific for this disease. Pleural effusion, either bilateral or unilateral, is found in slightly over one half of the patients (6, 9).

The electrocardiogram results are abnormal in 80% of the patients (6, 9). Classical electrocardiogram changes of pericarditis consist of ST elevation in two or all three standard limb leads, as well as in the precordial leads, without reciprocal ST depressions. The ST segments characteristically have

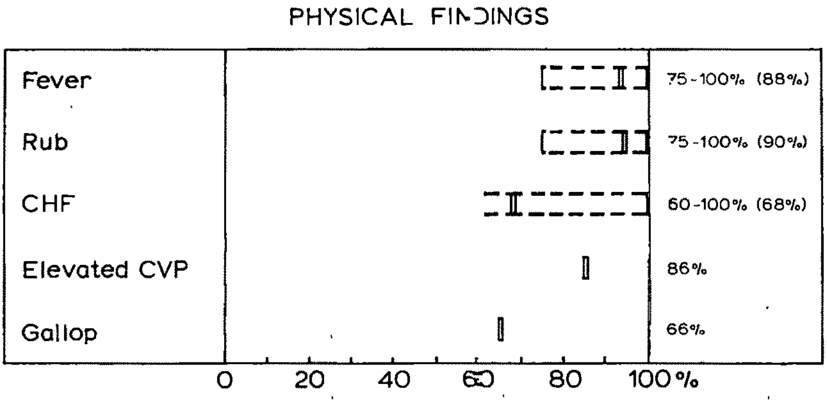


Figure 2 Physical findings on patients who developed pericarditis while on dialysis.

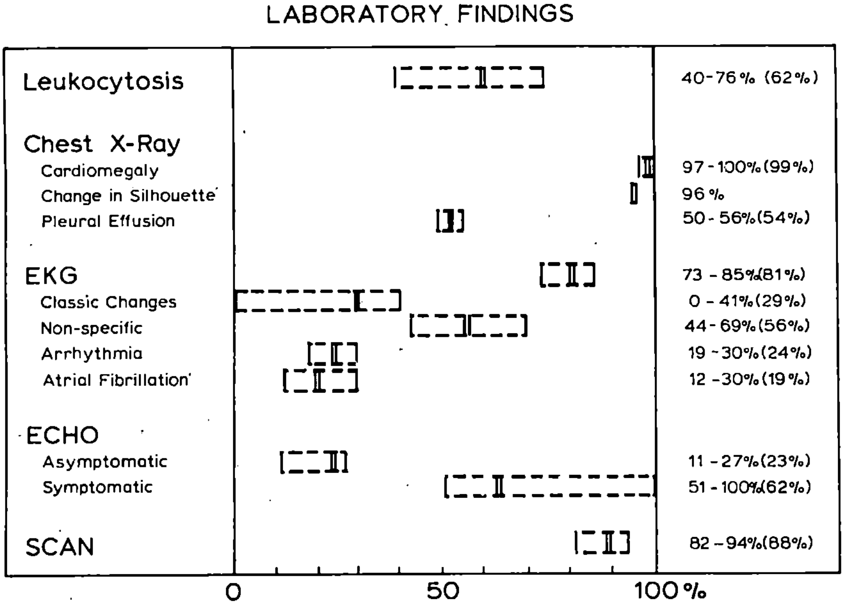


Figure 3 Laboratory findings in patients who develop pericarditis while on hemodialysis. The 10–30% incidence or frequency of positive echocardiography in asymptomatic patients refers to results of scanning all patients on dialysis. In any given dialysis population, approximately 25% will have signs of effusion on echocardiography. Most of this will vanish or remain stationary and only a few will progress to symptomatic pericardial effusion.

an upward concavity. Occasionally, ST depression may be seen in leads V1 and AVR. Over the next several days, the ST segments return to normal, then the T waves become inverted. Q waves do not develop. These classic findings are reported in about 30% of the patients with uremic pericarditis (6, 9, 24, 27).

Nonspecific changes that alter with therapy occur in 56% of the patients (6, 9, 27). Arrhythmias, atrial in origin, usually occur in about 25% (6, 9, 12), and atrial fibrillation occurs in one fifth of the patients (6, 25). These arrhythmias are sometimes resistant to therapy and may indicate underlying myocardial pathology (28).

Echocardiography is extremely helpful to the diagnosis of pericardial effusion. Abnormalities of myocardial function, chamber size, and valves can also be detected (29, 30). Pericardial effusions are seen as echo-free spaces between the epicardium and the posterior pericardium, or anterior to the right ventricular wall. It may be difficult to distinguish between a posterior pericardial effusion and a pleural effusion by echocardiography. The pericardial effusion should disappear behind the left atrium because of the pericardial reflection. Estimates of the volume of fluid present can be made (30), as well as possibly detecting the deposition of fibrin and loculation of fluid (29). Rarely, a swinging heart, which correlates with electrical alternans on electrocardiography, will be seen.

Effusions are found in about one fourth of the asymptomatic patients on dialysis and in two thirds of the patients with symptomatic pericardial disease (15–18, 24, 27). The absence of detectable effusion in some patients with pericarditis probably indicates an inflammatory noneffusive presentation of the disorder.

Asymptomatic effusions are usually small (16, 18), and frequently resolve spontaneously (15, 17), Although we classify this phenomenon with dialysis pericarditis, there is no evidence that it shares the same implications or pathophysiology (17). Of the patients on dialysis, 85–95% will at one time or another have some echocardiographic abnormality (15, 30), which includes left ventricular hypertrophy and dilatation, left atrial enlargement, hyperkinetic heart, decreased myocardial contractility, valvular abnormalities, and effusions.

Cardiac scan results are positive 88% of the time (6, 7, 27). The cardiac scan is superimposed over the chest roentgenogram to compare the outline with the silhouette.

COMPLICATIONS

The major complications of pericardial disease in renal failure include cardiac tamponade, constrictive pericarditis, and death (Figure 4). Some people suffer from a prolonged course of pericarditis.

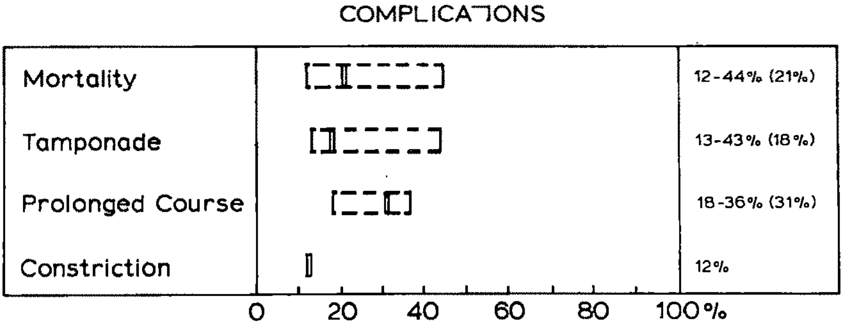


Figure 4 Complications in patients who develop clinically symptomatic pericarditis while on dialysis. Figures do not include complications in patients who have asymptomatic pericardial effusion detected by echocardiography.

Mortality

The mortality in patients who develop dialysis pericarditis is 21% (6, 9, 12, 24) and about 16% of those deaths are directly related to the pericarditis (6, 12). The long-term prognosis is probably worse in those patients who have developed pericarditis (14).

Death associated with uremic pericarditis is 48% (4, 5), which probably reflects the seriousness of the patients' condition when they begin hemodialysis. It may also be somewhat artificial because of recent improvements in our ability to treat uremia. As discussed later, uremic pericarditis appears to be quite responsive to dialysis therapy.

Tamponade

Pericardial tamponade is the most significant and most acutely life-threatening of the complications of pericarditis. Tamponade occurs in about 20% of the patients with dialysis pericarditis (6, 7, 9, 12, 24). This may be an underestimation of the threat of tamponade because adequate management of pericarditis may prevent the development of full-blown tamponade (12). Tamponade appears to be rare in patients who are asymptomatic, with small effusions diagnosed by echocardiography alone (17).

In general, the diagnosis of pericardial tamponade is clinical. The patient has hypotension with a narrow pulse pressure, elevated central venous pressure, and pulsus paradoxus. These hemodynamic alterations reverse after the removėl of pericardial fluid (31–33).

There are some unique aspects of tamponade physiology in uremia. The stress of dialysis frequently precipitates hypotension and what appears to be tamponade physiology (6, 24). A clue to the etiology of this phenomena is found in the report of Alfrey et al (33). They evaluated one patient who had a pericardial effusion with elevated right atrial pressure (RAP) and intrapericardial pressure (ICP). During dialysis the RAP and ICP fell, but not to normal. The cardiac index (CI) and stroke index (SI) also fell steadily —ultimately to 50% of their initial values. The femoral artery pressure (FAP) was maintained for almost four hours and then began to drop. At that point, the patient was hypotensive and had increased right-sided pressures, suggestive of tamponade. Pericardiocentesis was performed with a resultant fall in the RAP and ICP, and an increase in SI and CI, although not back to normal. The FAP did not rise initially. This suggests some degree of relative volume depletion accounting for the hypotension.

Alfrey and co-workers also noted minimal changes in the intrapericardial electrolytes or osmolality during dialysis, changes that were insufficient to account for tamponade during dialysis (33). Echocardiographic changes, however, show an increase in the effusion in patients on dialysis (24).

Another, perhaps unique, aspect of the hemodynamics of tamponade and renal disease, relates to pulsus paradoxus. Although thought to be almost universally present in tamponade (31), it is not consistently found in the clinical setting of renal failure (33, 34). In dogs it has been observed that as hypotension progresses, pulsus paradoxus may disappear (33). Patients with uremia who do not exhibit pulsus paradoxus, have evidence of significant underlying left ventricular disease (34). In either case, the absence of pulsus in a patient with other evidence of tamponade does not rule out tamponade physiology.

Other

Constrictive pericarditis is probably rare, although an incidence of 12% has been reported (6). A prolonged course of pyrexia and chest pain may occur in one third of the patients (6, 12).

PATHOLOGY AND FLUID ANALYSIS

The pathology of pericarditis associated with renal failure was reviewed by Baldwin & Edwards (35).

Uremic Pericarditis

The surfaces of the pericardium are diffusely involved with a fibrinous exudate (36), and fibrinous adhesions are almost always present (37). There is a layer of organizing perivascular connective tissue (36), and the pericardial space may be obliterated by fibrous tissue (36, 37). Inflammatory cells are almost all mononuclear, unless a concomitant bacterial infection is present (37). The myocardium may also be inflamed.

The pericardial fluid is usually serous, and may contain blood (23, 36). Over 50 years ago, hemorrhagic effusions accounted for 17% of the effusions found at autopsy (37). The volume of fluid was variable, ranging from zero to two liters (37).

Dialysis Pericarditis

The pathology of dialysis pericarditis is basically the same as uremic pericarditis, which probably reflects the limited response of the pericardium to stress. Thus, the fibrinous inflammatory reaction that involves both serosal membranes is seen. Vascular collagen tissue is laid down and adhesions may be formed. Hemorrhage is almost universal (9, 19, 38, 39).

The fluid is serosanguineous to bloody 97% of the time (12, 13, 19, 26, 33, 39, 40). The fluid hematocrit varies from less than 1% up to 29% (33, 40), and is almost always lower than the peripheral hematocrit (40). The level is stable with time, and there is no evidence that red cells enter the pericardial fluid during dialysis with regional heparinization (33).

The fluid is an exudate based on protein ratios with serum (33, 41). The white blood count varies from 450 to 7300. There may be neutrophilic predominance or there may be no particular cellular predominance. Blood urea nitrogen (BUN), creatinine (cr), and glucose do not vary significantly from serum values (33). Coagulation studies on hemorrhage effusions, not necessarily uremic, demonstrate increased fibrinogen, fibrin degradation products, and fibrinolytic activity (42). The volume of fluid varies from 50 ml to 3 liters (12, 38).

ETIOLOGY

The multiple etiologies of pericarditis were reviewed by Spodick (43). The dialysis patient is by no means protected from other causes of pericarditis, and during diagnosis other, especially the treatable, forms of pericarditis should be excluded first.

Uremic Pericarditis

Barach concluded in 1922 that uremic pericarditis was the result of azotemia (37). He noted that the average BUN was 269 and that the patients were uremic with frequent acidosis and a tendency to hemorrhage.

Several series demonstrate the role of azotemia in uremic pericarditis—the average BUN is 243 and cr is 18.2 (4, 5, 8, 36, 37). Although this is not significantly different from the markers of azotemia in patients without pericarditis (5, 8), there is a correlation between resolution of pericarditis and improvement of azotemia by dialysis (4, 5). Over-hydration may be a contributing factor, as one study showed a significant difference in weight gain between comparable patients with and without uremic pericarditis (11). Further evidence supporting the role of azotemia or fluid overload is the responsiveness of this form of pericarditis to dialysis.

There is no clear-cut etiologic role for infection (36, 37), and age and sex are not indicators of risk.

Dialysis Pericarditis

In most patients with dialysis pericarditis, a potential precipitant can be found. These patients all have azotemia and preceding infections and stress are common. Hyperuricemia and evidence of hyperparathyroidism may be present, and tamponade frequently occurs during or immediately after dialysis. These observations lead to several hypotheses regarding etiologic or exacerbating factors in uremic pericardial disease.

AZOTEMIA In patients with dialysis pericarditis, the BUN level averages 96 and the cr 11.9 (6, 7, 9, 13, 14), significantly lower than in uremic pericarditis. There is no difference, however, in average markers of azotemia between dialysis patients who develop pericarditis and those who do not (6, 7, 9, 11, 13, 24). Nerve conduction velocities are normal (6).

On the other hand, 30–50% of the patients who develop pericarditis have had previous difficulty with biochemical control, often related to shunt malfunction or home dialysis (6, 11). Bailey et al (4) noted that the incidence of dialysis pericarditis had decreased in their stable dialysis population since the frequency of dialysis has been increased to three times per week.

Peritoneal dialysis has a somewhat lower incidence of dialysis pericarditis than does hemodialysis. In a comparison of the two methods, the incidence of dialysis pericarditis was 4.3 per 100 patient years for peritoneal dialysis vs 8.4 per 100 patient years for hemodialysis. This occurred in spite of elevated markers of azotemia in the patients on peritoneal dialysis (14).

Thus, although we are discussing a disease of uremia, the direct contribution of azotemia is unclear. Certainly the uremic state predisposes to pericarditis and effusions, yet there must be other factors involved.

INFECTION Up to 90% of patients with pericardial disease in renal failure had preceding infections (4, 6, 9, 10, 13, 14, 25, 39) or a viral prodrome. Common sites of infections include dialysis vascular accesses, lungs, surgical wounds, and blood. On the other hand, direct microbiologic study results on pericardial effusions are almost always negative (9, 19) as are viral study results (6, 9, 25). A recent paper, however, based on serologic titers, has implicated a role for viruses in this syndrome (44). This would correlate with the high frequency of a viral prodrome (10, 13, 25).

On the other hand, of 55 patients with purulent pericarditis who were seen from 1943 to 1977 at Johns Hopkins Hospital, 9 (16%) had underlying chronic renal failure. This was the second most common underlying disease; 22% of these patients had another underlying source of infection including meningitis, pneumonia, and subacute bacterial endocarditis (45). Infection can only be directly implicated in a minority of cases of uremic pericarditis, but because of the ominous implications of purulent pericarditis, this possibility must be kept in mind. Clinical indicators of possible purulent pericarditis include a temperature in excess of 38.5°C, a white cell count of greater than 15,000, or an associated infection. Suspicion should be highest if the infection is either contiguous or hematogenous.

STRESS It is conceivable that the role of infections is mediated through secondary mechanisms. Thus, a previous infection could increase stress and catabolic activity, which could indirectly increase the risk of pericarditis.

Major surgery has frequently preceded pericarditis (6, 10, 13, 39), but Comty and co-workers (6) found only 2% of major operations on uremic patients were followed by pericarditis (6).

ANTICOAGULATION The presence of hemorrhagic fluid and the common occurrence of tamponade on dialysis suggest that anticoagulation may exacerbate the pericarditis and lead to hemorrhagic tamponade. Thus, in most hemodialysis centers, patients are regionally heparinized when pericarditis or effusion is noted (6, 7, 24).

There is little evidence that anticoagulation plays a significant role in dialysis pericarditis. Clinically, systemic heparinization does not appear to increase the incidence of pericardial tamponade (5, 9, 17). Effusions have been seen by echocardiography to increase in size under regional heparinization. The possible role of local interference with the clotting mechanism has been mentioned (42) as has an alternative hemodynamic explanation for acute decompensation on dialysis (33).

HYPERURICEMIA Hyperuricemia has also been implicated since uric acid concentrations are higher in patients with than in those without pericarditis (7, 24). Other series have failed to confirm this finding. The pericardial concentration of uric acid was reported to be higher than the serum concentration (13). Urate crystals have not been detected in the pericardial fluid or pericardium (7, 13).

HYPERPARATHYROIDISM Hyperparathyroidism is a common finding in patients with pericarditis (6, 11). In one series, 13% of all patients had hyperparathyroidism compared to 20% of the patients with pericarditis (6). Other studies have failed to support this etiology (7, 13, 39).

THERAPY

Uremic Pericarditis

Untreated, the prognosis is grim, with an 88% mortality within two days to two months (4), but uremic pericarditis responds to the initiation of dialysis therapy. Thus, 89% of 154 patients with uremic pericarditis responded to dialysis therapy, although the recurrence rate was 15% (4, 6–8, 11, 20).

Thus, about 25% of the patients do not respond to dialysis therapy or relapse. Most studies of therapy classify these patients as suffering from intractable pericarditis and do not separate them from patients with dialysis pericarditis. Their further therapy is discussed in the following section.

Dialysis Pericarditis

DIALYSIS Dialysis alone is much less successful for patients with dialysis pericarditis—only 41% of the patients respond (4, 6–8, 11–13) with a recurrence from 3–33% (11, 12). Only about one third of these patients will recover permanently, and 95% of those who respond do so within 10–14 days (19).

Peritoneal dialysis has been advocated over hemodialysis in the therapy of dialysis pericarditis (10). Charcoal hemoperfusion may also offer advantage over dialysis alone (46). Both of these approaches need further evaluation, although it seems unlikely that they will significantly expand the role of conservative medical therapy for this disorder.

SYSTEMIC ANTI-INFLAMMATORY AGENTS Systemic steroids do not significantly improve the results obtained with dialysis alone (6); indeed, possible steroid complications may be responsible for over 25% of the deaths (6). Indomethacin has been advocated for treatment (47), but the only controlled double-blinded study of this drug failed to show any beneficial effect (48).

PERICARDIOCENTESIS The role of pericardiocentesis alone is limited. Initial success in reducing the size of the effusion and in relieving hemodynamic compromise is probably near 80% (13, 21, 27, 40, 49). Recurrence occurs in about 70% of those who initially respond to treatment (7, 13, 21, 27, 40, 49), and the procedure carries an average mortality of 16% (7, 12, 13, 21, 26, 27, 40, 49). Nonfatal complications occur in 6–21% of the patients (19, 33, 40). Thus, satisfactory long-term results with this procedure are uncommon.

PERICARDIAL STEROIDS Buselmeier et al (41) combined the use of steroids with pericardiocentesis and improved the results in patients with intractable effusions. They combined indwelling pericardial catheter drainage and periodic instillation of triamcinolone and vastly improved the results over pericardiocentesis alone. This technique has been 97% successful in relieving effusion in a total of 57 patients (13, 41, 50, 51). Only one recurrence has been reported (41) and no deaths have occurred (13, 41, 50, 51). With wider dissemination of the approach, a 1–2% incidence of death would be expected, as seen in large series of pericardiocentesis (52). Modifications of this approach have included using only a single instillation of triamcinolone and an indwelling large-bore catheter (50, 51).

SURGICAL THERAPY Surgical approaches include drainage procedures such as pericardial window formation, pericardiostomy, and pericardiectomy. The drainage procedures are successful 94% of the time although these results are marred by recurrence and mortality rates of 6% each and a 4% complication rate (7, 11, 12, 38). Complications include infection and early closure of the drainage site. Thus, approximately 20% of patients undergoing these procedures have an unsatisfactory outcome.

With a success rate of 97%, pericardiectomy has been the standard treatment for intractable effusions. It has a significant complication rate of only 3% and a mortality rate of 1% (7, 19, 21, 27, 38, 39, 49). Recurrences are rare but when they occur, they may be related to an incomplete removal of the pericardium (19) (Table 2).

OUR APPROACH Our approach to pericarditis and pericardial effusion in renal failure is as follows. If the patient has uremic pericarditis, or dialysis pericarditis preceded by a period of poor biochemical control, we initiate daily six-hour dialysis with regional heparinization. We would expect a therapeutic response within 10–14 days. If we do not see a change in the size of the effusion over two weeks, we consider the effusion intractable.

If the patient has dialysis pericarditis with an effusion and evidence of hemodynamic compromise, we believe that a definitive procedure should be performed. Likewise, if the patient develops tamponade, that requires emergent action, or if the patient has an intractable effusion, we recommend a definitive procedure.

For reasons we have reviewed (Table 2), we do not classify pericardiocentesis as a definitive procedure. If it must be done emergently, at a minimum, a catheter should be left in place.

It is difficult to define a clinical difference in outcome between patients treated by pericardiectomy and those treated by intrapericardial steroids (Table 2). The results in both cases depend upon the comparative quality of the techniques as well as the skill of the operators. We prefer Buselmeier's approach. It is less invasive, is as successful and safe, and requires a shorter period of hospitalization (41).

CONCLUSIONS

Our understanding of the range of disorders related to pericarditis in end-stage renal disease has evolved significantly over the past decade. It is clear that there are at least two clinically different entities associated with this phenomenon. Uremic pericarditis tends to be a clinically milder disease. It is more likely associated etiologically with the accompanying azotemia and fluid overload. It is reasonably responsive to the initiation of dialysis therapy.

Table 2 Results of various treatments for symptomatic pericarditis occurring in dialyzed patients

Therapy	No.	Failed (%)	Recurrence (%)	Complications (%)	Death (%)
Pericardiocentesis	63	20	70	6–21	16
Pericardial steroids	59	3	2	8	—
Pericardial window	53	6	6	4	—
Pericardiectomy	91	3	≤1	3	1

Dialysis pericarditis is more frequently associated with systemic symptoms and signs. Its etiology is unclear. It is more likely to require invasion of the pericardial space to achieve lasting recovery.

The management of uremic pericarditis requires a substantial amount of judgement by the physician, and meticulous monitoring of the patient. We have outlined our therapeutic approach. It is important to keep in mind that the pericarditis may have a cause other than renal failure. Appropriate diagnostic studies should be carried out, especially if infective pericarditis is suspected.

Because of the risk of tamponade, all these patients should be monitored in an intensive care unit. Chest roentgenograms, echocardiography, and careful monitoring of hemodynamic data should be obtained on a regular basis.

If symptoms are only mild and no serious hemodynamic problems occur, a ten-day treatment with intensified dialysis can resolve approximately one half of these cases. If the patient fails to respond within ten days or develops increased pericardial effusion as seen on roentgenogram, echo, or scan, or becomes hemodynamically unstable, we advocate pericardiocentesis, either blind or under direct vision, followed by an indwelling catheter and periodic instillation of steroids. Some people prefer to perform pericardiectomy but this procedure is much more invasive, requires much more prolonged hospitalization, is not more successful, and has at least as many complications. We believe it to be inferior to the therapy described by Buselmeier and his associates.

With present technology and access to medical care, we should be able to achieve a mortality rate of dialysis pericarditis, uncomplicated by other life-threatening disorders, of less than 1%.

Literature Cited

1. Bright, R. 1836. Tabular view of the morbid appearances in 100 cases connected with albuminous urine: with observations. *Guy Hosp. Rep.* 1:380–400
2. Ogburn, H. M., Ayrus, J. C., Eknoyan, G. 1978. Pericarditis in renal failure; therapeutic considerations. *Dialysis Transplant.* 7:1133–38
3. Comty, C. M., Wathen, R. L., Shapiro, F. L. 1976. Uremic pericarditis. *Cardiovasc. Clin.* 7:219–35
4. Bailey, G. L., Hampers, C. L., Hager, E. B., Merrill, J. P. 1968. Uremic pericarditis; clinical features and management. *Circulation* 38:582–91
5. Skov, P. E., Hansen, H. E., Spencer, E. S. 1969. Uremic pericarditis. *Acta Med. Scand.* 186:421–28
6. Comty, C. M., Cohen, S. L., Shapiro, F. L. 1971. Pericarditis in chronic uremia and its sequels *Ann. Intern. Med.* 75:173–83
7. Marini, P., Hull, A. R. 1975. Uremic pericarditis: a review of incidence and management. *Kidney Int.* Suppl. 2, pp. 163–68
8. Wray, T. M., Stone, W. J. 1976. Uremic pericarditis: a prospective echocardiographic and clinical study. *Clin. Nephrol.* 6:295–302
9. Mitchell, A. G. 1974. Pericarditis during chronic haemodialysis therapy. *Postgrad. Med. J.* 50:741–45
10. Kramer, P., Wigger, W., Scheler, F. 1975. Management of uraemic pericarditis. *Br. Med. J.* 2:564–66
11. Jungers, P., Zingraff, J., Brefort, G., Moynot, A., Drueke, T., Man, N. K., Crosnier, J. 1978. Uremic pericarditis: beneficial effect of surgical pericardiostomy. *Kidney Int.* 14:202 (Abstr.)
12. Ribot, S., Frankel, H. J., Gielchinsky I., Gilbert, L. 1974. Treatment of uremic pericarditis. *Clin. Nephrol.* 2:127–30
13. Fuller, T. J., Knochel, J. P., Brennan, J. P., Fetner, C. D., White, M. G. 1976. Reversal of intractable uremic pericarditis by triamcinolone hexacetonide. *Arch. Intern. Med.* 136:979–82
14. Silverberg, S., Oreopoulos, D. G., Wise, D. J., Uden, D. E., Meindok, H., Jones, M., Rapoport, A., deVeber, G. A. 1977. Pericarditis in patients undergoing long-term hemodialysis and peritoneal dialysis. *Am. J. Med.* 63:874–80
15. Horton, J. D., Gelfand, M. C., Sherber, H. S. 1977. Natural history of asymptomatic pericardial effusions in patients on maintenance hemodialysis. *Proc. Dialysis Transplant. Forum* 7:76–81
16. Kleiman, J. H., Motta, J., London, E., Pennel, J. P., Popp, R. L. 1978. Pericardial effusions in patients with end-stage renal disease. *Br. Heart J.* 40:190–93
17. Goldstein, D. H., Nagar, C., Srivastava, N., Schacht, R. A., Ferris, F. Z., Flowers, N. C. 1977. Clinically silent pericardial effusions in patients on long-term hemodialysis. *Chest* 72:744–47
18. Lazarus, J. M., Gottlieb, M. N., Lowrie, E. G., Techoltz, L., Merrill, J. P. 1976. Echocardiographic findings in stable hemodialysis patients. *Proc. Dialysis Transplant. Forum* 6:53–57
19. Connors, J. P., Kleiger, R. E., Shaw, R. C., Voiles, J. D., Clark, R. E., Hartner, H., Roper, C. L. 1976. The indications for pericardiectomy in the uremic pericardial effusion. *Surgery* 80:689–94
20. El-Said, W., Gabal, I. A. 1978. Treatment of uremic pericarditis and pericardial effusion by augmented hemodialysis. *Int. Urol. Nephrol.* 10:53–58
21. Hollomby, D., Gonda, A., Morin, J., Long, R., Dobell, A. R. C. 1975. Uremic pericardial tamponade. *Proc. Dialysis Transplant. Forum* 5:16–19
22. Guild, W. R., Bray, G., Merrill, J. P. 1957. Hemopericardium with cardiac tamponade in chronic uremia. *N. Engl. J. Med.* 257:230–31
23. Hagar, E. B. 1965. Clinical observations on five patients with uremic pericardial tamponade. *N. Engl. J. Med.* 273:304–8
24. Winney, R. J., Wright, N., Sumerling, M. D., Lambie, A. T. 1977. Echocardiography in uraemic pericarditis with effusion. *Nephron* 18:201–7
25. Buselmeier, T. J., Simmons, R. L., Najarian, J. S., Mauer, S. M., Matas, A. J., Kjellstrand, C. M. 1976. Uremic pericardial effusion. *Nephron* 16: 271–80
26. Singh, S., Newmark, K., Ishikawa, I., Mitra, S., Berman, L. B. 1974. Pericardiectomy in uremia. The treatment of choice for cardiac tamponade in chronic renal failure. *J. Am. Med. Assoc.* 228:1132–35
27. Morin, J. E., Mulder, D. S., Long, R. 1976. Pericardiectomy for uremic tamponade. *Can. J. Surg.* 19:109–12
28. Spodick, D. H. 1976. Arrhythmias during acute pericarditis. A prospective study of 100 consecutive cases. *J. Am. Med. Assoc.* 235:39–41
29. D'Cruz, I. A., Bhatt, G. R., Cohen, H. C., Glick, G. 1978. Echocardiographic detection of cardiac involvement in patients with chronic renal failure. *Arch. Intern. Med.* 138:720–24

30. Schott, C. R., LeSar, J. F., Kotler, M. N., Parry, W. R., Segal, B. L. 1978. The spectrum of echocardiographic findings in chronic renal failure. *Cardiovasc. Med.* Feb., pp. 217–27
31. Hancock, E. W. 1979. Cardiac tamponade. *Med. Clin. North Am.* 63: 223–37
32. Shabetai, R., Fowler, N. O., Guntheroth, W. G. 1970. The hemodynamics of cardiac tamponade and constrictive pericarditis. *Am. J. Cardiol.* 26:480–88
33. Alfrey, A. C., Goss, J. E., Ogden, D. A., Vogel, J. H. K., Holmes, J. H. 1968. Uremic hemopericardium. *Am. J. Med.* 45:391–400
34. Reddy, P. S., Curtiss, E. I., O'Toole, J. D., Shaver, J. A. 1978. Cardiac tamponade: hemodynamic observations in man. *Circulation* 58:265–72
35. Baldwin, J. J., Edwards, J. E. 1976. Uremic pericarditis as a cause of cardiac tamponade. *Circulation* 53:896–901
36. Buja, L. M., Friedman, C. A., Roberts, W. C. 1970. Hemorrhagic pericarditis in uremia. Clinicopathologic studies in six patients. *Arch. Pathol.* 90:325–30
37. Barach, A. L. 1922. Pericarditis in chronic nephritis. *Am. J. Med. Sci.* 163:44–60
38. Engelman, R. M., Levitsky, S., Konchigeri, H. N., Wyndham, C. R. C., Roper, K., Kurtzman, N. A. 1977. Total pericardiectomy for uremic pericarditis. *World J. Surg.* 1:769–76
39. Ali-Regiaba, S., Gay, W. A., Sullivan, J. F., Tapia, L., David, D. S., White, R. P., Stenzel, K. H., Riggio, R. R., Cheigh, J. S., Rubin, A. L. 1974. Treatment of uraemic pericarditis by anterior pericardiectomy. *Lancet* 2:12–14
40. Goldberg, M., Lazarus, J. M., Gottlieb, M. N., Lowrie, E. G., Merrill, J. P. 1975. Treatment of uremic pericardial effusion. *Proc. Dialysis Transplant. Forum* 5:20–25
41. Buselmeier, T. J., Davin, T. D., Simmons, R. L., Najarian, J. S., Kjellstrand, C. M. 1978. Treatment of intractable uremic pericardial effusion. Avoidance of pericardiectomy with local steroid instillation. *J. Am. Med. Assoc.* 240:1358–59
42. Kluge, T., Hall, K. V. 1976. Surgery in acute and chronic pericarditis. Pathophysiology and management. *Scand. J. Thorac. Cardiovasc. Surg. 10:21–30*
43. Spodick, D. H. 1971. Differential diagnosis of acute pericarditis. *Prog. Cardiovasc. Dis.* 14:192–203
44. Osanloo, E., Shalhoub, R. J., Cioffi, R. F., Parker, R. H. 1979. Viral pericarditis in patients receiving hemodialysis. *Arch. Intern. Med.* 139:301–3
45. Klacsmann, P. G., Bulkley, B. H., Hutchins, G. M. 1977. The changed spectrum of purulent pericarditis. An 86 year autopsy experience in 200 patients. *Am. J. Med.* 63:666–73
46. Martin, A. M., Gibbins, J. K., Kimmitt, J., Rennie, F. 1979. Hemodialysis and hemoperfusion in the treatment of uremic pericarditis. A study of 13 cases. *Dialysis Transplant.* 8:135–62
47. Minuth, A. N. W., Nottebohm, G. A., Eknoyan, G., Suki, W. N. 1975. Indomethacin treatment of pericarditis in chronic hemodialysis patients. *Arch. Intern. Med.* 135:807–10
48. Spector, D., Alfred, H., Siedlecki, M., Briefel, G. 1978. Indomethacin treatment of uremic pericarditis. *Am. Soc. Nephrol.,* p. 30A (Abstr.)
49. Morin, J. E., Hollomby, D., Gonda, A., Long, R., Dobell, A. R. C. 1976. Management of uremic pericarditis: a report of 11 patients with cardiac tamponade and a review of the literature. *Ann. Thorac. Surg.* 22:588–92
50. Oakes, D. D., Weidig, J. C., Spees, E. K., Light, J. A. 1977. Another view on treatment for uremic pericardial effusion (Letter). *Surgery* 82:894
51. Popli, S., Ing, T. S., Daugirdas, J. T., Stroka, G. S., Osanloo, E. O., Vilbar, R. M., Geis, W. P. 1979. Treatment of uremic pericardial effusion by local steroid instillation via subxiphoid pericardiotomy. *J. Dialysis* 3:
52. Krikorian, J. G., Hancock, E. W. 1978. Pericardiocentesis. *Am. J. Med.* 65: 808–14

Ann. Rev. Med. 1980. 31:361–71
Copyright © 1980 by Annual Reviews Inc. All rights reserved

THE ENDOTHELIUM AND THROMBOSIS[1]

❖7366

Robert T. Wall, M.D.

Department of Medicine, Division of Hematology, Stanford Medical School,
Stanford, California 94305

Laurence A. Harker, M.D.

Department of Medicine, Division of Hematology, University of Washington
School of Medicine, Seattle, Washington 98195

The vascular endothelium is a metabolically active tissue of mesodermal origin at the interface between blood and the underlying vascular structures. Endothelial cells have a number of critical functions that include (*a*) the material transfer of metabolic substances of varied molecular size between the circulating blood and the surrounding tissues, (*b*) the formation of a relative barrier to blood cells, macromolecular and particulate material, (*c*) the creation of a synthetic and metabolic interface between blood regulatory factors affecting the function of the vascular wall and vascular factors that affect the function of blood components, (*d*) the maintenance of thromboresistance, and (*e*) the mediation of vascular repair processes of cell migration, proliferation, and thrombolysis (1–5).

The endothelium consists of a stable population of polygonally shaped cells forming a continuous monolayer on the vascular luminal surface. The endothelium is not homogeneous: (*a*) aortic endothelial cells are generally thicker than those of capillaries and veins, (*b*) the endothelium of visceral capillaries is remarkable for its fenestrations, and (*c*) intercellular junctions progress from the least complex junctions of the venules to the elaborate organization of arterioles, which contain both tight junctions and gap junc-

---

[1] This work was supported in part by grants HL-11775 and HL-18645; and by Young Investigators Research Award HL-24117 (Dr. Wall) from the US Public Health Service.

0066-4219/80/0401-0361$01.00

tions (6, 7). It is likely that this structural variation is related to functional requirements imposed by blood flow and the characteristics of the tissues served. The exchange of materials across the endothelium involves movement by active vesicles, transendothelial channels, or intercellular clefts. Maintaining the integrity of the vessel wall to the egress of blood cells depends upon the barrier function of the endothelium and connective tissue vascular components, together with the presence of functional platelets which seal over gaps that develop in the endothelial lining. In addition, platelets may enhance the integrity and continuity of vascular endothelium by some other mechanism (8).

The endothelium contributes to the generation of its underlying connective tissue, which is composed primarily of several classes of collagen, glycosaminoglycans, elastin, and microfibrils (9, 10). This subendothelial structure varies from a well-defined basement membrane in capillaries to a thin, less-well-defined area in the aorta. This connective tissue matrix regulates to some extent the permeability of the inner vessel wall and provides the principal stimulus to thrombosis following vessel injury (11, 12). Defects in the endothelium lead to thrombosis, the nature of which will depend on the locus, extent, duration, and type of endothelial lesion. Endothelial disruption directly activates all components of the hemostatic apparatus: (*a*) The vessel constricts rapidly. (*b*) Platelets adhere immediately to exposed subendothelial connective tissue structures, particularly collagen fibers. Adherent platelets recruit additional platelets by releasing platelet constituents, including ADP and the products of arachidonate metabolism synthesized during the secretory process. Adherent and aggregated platelets enhance vasoconstriction by releasing thromboxane A_2 and vasoactive amines. (*c*) Coagulation is initiated through the intrinsic system by the activating effect of collagen and elastin on factor XII, and through the extrinsic system by tissue factor activation of factor VII to form fibrin. (*d*) Fibrinolysis follows the release of plasminogen activators from damaged vascular intima (13–16).

THROMBORESISTANCE

One of the basic functional characteristics of intact, normal endothelium is its nonreactivity to platelets, leukocytes, or coagulation systems. The surface properties conferring this thromboresistant state may be a consequence of certain cell surface glycoproteins, the so-called glycocalyx, or the surface charge produced by these cell surface components. Moncada et al (17) described the synthesis of prostaglandin I_2 (PGI_2) by blood vessel rings and observed that this prostaglandin inhibited platelet aggregation and, at high concentrations, platelet adhesion. Moreover, modulation of PGI_2 production by injury factors including activated clotting enzymes serves to limit locally any hemostatic response (18). The capacity of the endothelial lining to regulate PGI_2 production is thought to be one mechanism contributing to the nonthrombogenic properties of intact vascular endothelium. It has not yet been shown that PGI_2 synthesis plays an important role in maintaining endothelial thromboresistance under physiological condition. Endothelial desquamation results in a loss of PGI_2-synthesizing surface as well as exposure of subendothelial connective tissue structures; thrombus formation results.

The recently developed technology for the in vitro study of the biologic properties of vessel wall cells in isolation has greatly contributed to our understanding of the interaction of platelets, plasma coagulation factors, and the endothelium (5, 19). Cultured endothelium, but also cultured vascular smooth muscle cells and to a lesser degree fibroblasts, can be stimulated to synthesize PGI_2 from endogenous arachidonate (20, 21). The endothelium apparently does not utilize exogenous PGI_2 precursors such as might be provided by the platelet unless platelet thromboxane production is inhibited (22, 23).

PGI_2 production is sensitive to inhibitors of the enzyme cyclooxygenase such as aspirin. Thus, aspirin, an inhibitor of platelet function, might paradoxically enhance thrombosis by inhibiting endothelial cell PGI_2. Previous studies employing aspirin as an antithrombotic agent may not have been ideal because the aspirin concentrations achieved might have inhibited endothelial cell PGI_2 production (24). In this regard, specific inhibitors of platelet thromboxane synthetase have been postulated to have an additive antithrombotic effect by providing platelet endoperoxides to vessel wall cells for possible PGI_2 synthesis (25). PGI_2 also relaxes vascular smooth muscle (26). While it is possible that a relative reduction in PGI_2 might contribute to arterial spasm and thereby contribute to coronary and carotid insufficiency syndromes, there is little evidence to date that PGI_2 circulates in the resting, nonstimulated state.

The interaction of coagulation factors with the vessel wall has implications for the regulation of hemostasis and thrombosis. Thrombin binds to cultured endothelial cells (27) and is reported to increase the production of PGI_2 (20). Somewhat paradoxically, this binding of thrombin leads to increased platelet adherence to the endothelium in vitro (28).

Factor VIII:von Willebrand factor is associated with the endothelium in vivo and in vitro (29, 30). The association of this protein with the endothelium is of interest since Factor VIII:von Willebrand factor is necessary for platelet adhesion to vascular subendothelium in high shear systems, such as the capillaries and arteries (31). Endothelial cell synthesis of this protein has been reported (30). Recent studies suggest that the association of Factor VIII and endothelial cells is at least in part also due to binding of the protein to the endothelial cell surface (32). The bound Factor VIII:antigen can be released by β-adrenergic stimulation, and is probably available for mediating platelet adhesion at sites of intimal damage.

ARTERIAL VASCULAR DISEASE

Fibromuscular intimal plaques are induced in experimental animals by mechanical, chemical, and immune endothelial cell injury (33–35). The resultant lesions are similar to early atherosclerotic lesions in man. Intimal lesion formation following experimental endothelial cell injury is dependent on platelets (36, 37). Platelets are now known to contain a growth factor in their α-granules that is necessary for smooth muscle cell proliferation in vitro (38–40).

Recent studies suggest that the amount of endothelial cell desquamation determines the rate of platelet consumption (Table 1), and in turn the severity of the intimal smooth muscle cell lesion (34, 41). Repair of the endothelial lesion occurs by migration of adjacent intact cells into the lesion and subsequent generation of new cells behind the leading edge (42–44). Preliminary evidence suggests that this process is followed by regression or lack of progression of the smooth muscle cell lesion (45, 46).

Examination of the aortae of experimental animals clearly demonstrates focal areas of increased endothelial cell turnover in the normal animal (47,

Table 1 Platelet consumption by the aortic subendothelium of monkeys[a]

	Length of deendothelialized aortic segment (cm)		
Platelet	0	10	20
Recovery following balloon catheterization (%)	85 ± 4	69 ± 5[b]	53 ± 5[b]
Survival (days)			
0 (injury) week	8.0 ± 0.6	5.4 ± 0.5[b]	3.3 ± 0.4[b]
1 week	8.1 ± 0.5	6.6 ± 0.6[b]	4.9 ± 0.5[b]
2 weeks	7.9 ± 0.6	7.8 ± 0.5	6.1 ± 0.6[b]
3 weeks	8.0 ± 0.5	8.0 ± 0.5	7.7 ± 0.5

[a] Aortic deendothelialization was produced in monkeys (*Macaca nemestrina*) by balloon catheter desquamation of 10 or 20 cm segment of aorta. The animals' platelets were previously labeled with ^{51}Cr. Sixteen and 32% of the circulating ^{51}Cr-platelets disappeared immediately following the removal of the endothelium from a 10 and 20 cm segment, respectively. Similarly platelet survival was reduced by an amount that was related to the area of exposed subendothelium. Catheterized control animals are shown in the "0" desquamation column. Normalization of platelet survival occurred over two to three weeks. Each value represents the mean (± 1 SD) of at least six animals.

[b] $P < 0.01$

48). These focal areas occur throughout the aorta but are increased at vessel branch points and are associated with increased permeability as evidenced by vital dye staining. At such sites of increased cell turnover and decreased cell density, endothelial cells appear to internalize more surface-bound material, such as low density lipoprotein cholesterol, than occurs at confluent dense sites (49, 50). This mechanism might account for the increased permeability at these sites.

We recently described a platelet constituent that stimulates the migration but not the proliferation of endothelium in vitro, but both the migration and proliferation of smooth muscle cells (51). The chemical identity of this factor is not yet known. We hypothesize that this factor induces the migration of medial smooth muscle cells into the intima following endothelial cell desquamation and perhaps contributes to the endothelial repair process. In this regard, macrophages secrete a factor(s) that stimulates the preconfluent growth of cultured endothelial cells (52). Endothelial cells in culture release a factor(s) that can enhance smooth muscle cell proliferation (53, 54). The physiologic role of this factor has not been established.

We have examined the potential for using hyperlipidemia-induced alterations in the endothelium in vitro to explore the mechanism underlying the relationships between hypercholesterolemia and atherosclerosis. In an agarose migration assay, serum and plasma from individuals homozygous and heterozygous for familial hypercholesterolemia markedly inhibited endothelial cell migration (Table 2) without evidence of cell injury as measured in vitro by ^{51}Cr release and detachment assays (55). We postulate that inhibition of migration may contribute to the decreased endothelial cover-

Table 2 Effect of familial hypercholesterolemic sera on endothelial cell de novo surface coverage

		Whole blood serum		Cell-free plasma-derived serum		Whole blood serum/ normal cell-free plasma-derived serum[a]	
	Chol/tri (mg/dl)	Area covered		Area covered		Area covered	
		(mm^2)	Percent	(mm^2)	Percent	(mm^2)	Percent
Control	(189/125)	37.2 ± 4.3	100	22.1 ± 1.3	100	30.3 ± 2.0	100
Patient A	(652/112)	18.4 ± 0.9	49[b]	12.5 ± 1.9	57[b]	24.9 ± 2.2	85[c]
Patient L	(520/117)	19.2 ± 3.1	52[b]	10.9 ± 1.1	49[b]	21.7 ± 3.4	72[d]
Heterozygote pool	(382/134)	21.0 ± 5.2	57[b]				

[a] A mixing experiment with cells plated in normal or patient whole blood serum but normal cell-free plasma-derived serum is used to constitute the agarose gel to test for correction by the normal serum. Variance about the mean is ± 1 SD.

[b] $P < 0.0001$.

[c] $P < 0.01$.

[d] $P < 0.001$.

age either during growth of the artery or following endothelial cell loss. Serum from patients with mixed hyperlipidemia and hypertriglyceridemia had no effect in these assays.

Of potential significance in the pathogenesis of arterial vascular disease is the observation that sulfhydryl compounds at a concentration of 0.1 mM cause the desquamation of endothelial cells from a connective tissue substrate in vivo and from plastic in vitro. Higher concentrations of these agents result in endothelial cell disruption. By comparison smooth muscle cells are relatively resistant to these agents (56). Homocystinuria is characterized by elevations of the sulfur-containing amino acids homocysteine, homocystine, and methionine; it is also associated with thrombotic vascular disease and premature atherosclerosis. Similar lesions can be induced in primates by the continuous intravenous infusion of homocysteine (34). Since homocysteine is a sulfhydryl-containing compound, it may produce focal endothelial cell loss in vivo, allowing for platelet-mediated smooth muscle cell migration and proliferation. Moreover mixed disulfides are found in the plasma of individuals with accelerated atherosclerosis in the absence of cystathionine synthetase deficiency, which suggests that an increased production of sulfhydryl compounds in these patients may contribute to their atherosclerotic process (57).

ACUTE SMALL VESSEL THROMBOSIS

Small vessel thrombosis is primarily a platelet-consumptive event (58) resulting in thrombocytopenia, red cell fragmentation and organ dysfunction. Disorders such as thrombotic thrombocytopenic purpura (TTP), the hemolytic-uremic syndrome, malignant hypertension, eclampsia, and some forms of collagen vascular disease are characterized by evidence of endothelial cell injury, subendothelial deposits of a hyaline-like material composed of platelet and plasma proteins, and occlusive thrombi (59). Some of these disorders might result from acute widespread endothelial cell injury. To study this possibility directly, we examined the serum and plasma from seven individuals with classic TTP by using an endothelial cell ^{51}Cr release assay. Heat-inactivated sera were not cytotoxic to endothelium. However, sera from six of the seven patients produced significant specific release following the addition of fresh complement to the heat-inactivated serum. The sera were generally positive against multiple single-donor endothelial cell sources. Since six of the patients had also received at least some transfused blood products within a week to 10 days prior to obtaining the test serum, we could not exclude the possibility of alloantibody-mediated cytoxicity.

Subsequently, two patients with TTP who had never been transfused or pregnant were studied. The serum from one of these patients demonstrated cytotoxicity, while the second patient's serum was negative when tested.

The positive assay became negative during remission in the one long-term survivor from whom such a sample could be obtained. The serum factor could be absorbed by unrelated cultured endothelial cells and was not cytotoxic for cultured vascular smooth muscle cells. The separated lymphocytes of these patients were not cytotoxic for endothelial cells tested (60).

These preliminary observations, the possible clinical benefit of platelet-inhibitory agents in the treatment of this disease (61), and perhaps more important the excellent response of TTP patients to intensive plasmapheresis with 5% serum albumin replacement (62), suggest to us that at least some patients with TTP may have a process of autoimmune endothelial cell injury going on in the microvasculature. Indeed, acute endothelial cell injury may be involved in the development of a wide variety of small vessel diseases.

It has been suggested recently that TTP is a consequence of a plasma factor deficiency leading to unrestrained platelet aggregation. The corollary of this hypothesis is that TTP can be treated with large volume plasma infusion (63). Unfortunately results with such therapy are not uniformly successful, and the putative factor has not been identified (64). We suggest, therefore, that TTP be treated initially with intensive plasmapheresis using 5% albumin replacement together with antiplatelet agents. Blood component therapy should be given for replacement of depleted clotting factors and blood elements lost during the process of plasmapheresis. Clearly, the underlying mechanism(s) of TTP has not been definitively demonstrated.

Endothelial cells may play an important role in the clearance of immune complexes associated with various vasculitis syndromes (65) through a local stimulus to complement activation and neutrophil response. We have been unable to find C_{3b} and Fc receptors on freshly isolated endothelial cells, but other factors such as blood flow, complex size, or the nature of the antigen may be considerations.

Injection of endotoxin into experimental animals reportedly produces an increased endothelial cell turnover in large vessels and thrombosis of the microvasculature (66–68). Endothelial cell injury in this situation could release tissue factor and initiate the extrinsic clotting mechanism, as well as expose the subendothelium to plasma and activate the intrinsic system. Platelet thrombosis at such sites might also occur. In an interesting study of the Shwartzman phenomenon, infused endotoxin appeared to bind to the blood vessel wall and to lead to vessel injury and thrombosis when the vessel was reperfused with whole blood (69). The roles of lipopolysaccharide, blood neutrophils, and complement in endotoxin-dependent thrombotic disease remain to be defined. The use of anti-inflammatory agents in the prevention of these endotoxin-dependent events is currently under active investigation.

VENOUS THROMBOSIS

The mechanisms underlying venous thrombosis are less well defined than those underlying arterial thrombosis. Both platelets and coagulation factors are consumed during the formation of venous thrombosis in contrast to selective platelet utilization in arterial and microvascular thrombosis (70). The effectiveness of platelet-inhibitory agents in arterial thrombotic disorders, including prosthetic grafts and perhaps coronary and carotid disease, and the effectiveness of heparin in venous disease, are consistent with the kinetic data (71, 72). The initial events in venous thrombosis are poorly understood and the role of endothelial cell injury is speculative. Stewart and co-workers (73) observed increased neutrophil sticking to venous endothelium in operated animals followed by evidence of endothelial disruption. In another study (74) continuous infusion of lidocaine, to inhibit neutrophil function, effectively decreased the incidence of post-operative venous thrombosis in humans. These results suggest a role for inflammatory endothelial cell injury in the genesis of some episodes of venous thrombosis. However, direct trauma and stasis to the venous system appear to predominate in the genesis of most venous thrombotic events.

CONCLUSION

Endothelial cell injury may be a common mechanism in the development of many forms of acute and chronic vascular disease. At present, the evidence for a pathogenetic role of endothelial injury is primarily based on experimental models, but the opportunity to study endothelial cells in culture allows for additional direct testing of this hypothesis. Therapies implied by this postulate include measures to prevent endothelial injury, to protect the endothelium from injury, to modify neutrophil interaction with the vessel wall, and to inhibit platelet function. An example of such an experimental approach is the attempt to modify homocysteine-induced experimental atherosclerosis. Dipyridimole, an agent that inhibits platelet adhesion, prevented lesion formation without any measurable effect on endothelial cell desquamation. Sulfinpyrazone also prevented the smooth muscle cell lesion but appeared to act at least in part by protecting endothelial cells from the desquamative effects of injury (75).

Literature Cited

1. Majno, G., Joris. I. 1978. Endothelium 1977: A review. *Adv. Exp. Med. Biol.* 104:169–225
2. Chien, S. 1978. Transport across arterial endothelium. *Prog. Hemostasis Thromb.* 4:1–36
3. Cliff, W. J. 1976. The endothelium. *Blood Vessels,* pp. 31–67. Cambridge Univ. Press
4. Gingrich, R. D., Hoak, J. C. 1979. Platelet-endothelial cell interactions. *Semin. Hematol.* 16:208–20
5. Gimbrone, M. A. Jr. 1976. *Prog. Hemostasis Thromb.* 3:1–28
6. Simionescu, N., Simionescu, M., Palade, G. E. 1976. Structural-functional correlates in the transendothelial change of water-soluble macromolecules. *Thromb. Res.* 8:(Suppl II) 257–69
7. Simionescu, N., Simionescu, M., Palade, G. E. 1976. Recent studies on vascular endothelium. *Ann. NY Acad. Sci.* 275:64–75
8. Kitchens, C. S., Weiss L. 1975. Ultrastructural changes of endothelium associated with thrombocytopenia. *Blood* 46:567–78
9. Jaffe, E. A., Minick, C. R., Adelman, B., Becker, C. G., Nachman, R. 1976. Synthesis of basement membrane collagen by cultured human endothelial cells. *J. Exp. Med.* 144(1):209–25
10. Howard, B. V., Macarak, E. J., Gunson, D., Kelfalides, N. A. 1976. Characterization of the collagen synthesized by endothelial cells in culture. *Proc. Natl. Acad. Sci. USA* 713:2361–64
11. Karnovsky, M. J. 1979. The ultrastructure of glomerular filtration. *Ann. Rev. Med.* 30:213–24
12. Thorgeirsson, G., Robertson, A. L. Jr. 1978. The vascular endothelium—Pathobiologic significance. *Am. J. Pathol.* 93:803–48
13. Baumgartner, H. R., Muggli, R., Tschopp, T. B., et al. 1976. Platelet adhesion, release and aggregation in flowing blood. Effects of surface properties and platelet function. *Thromb. Haemostasis* 35:124–38
14. Burch, J. W., Majerus, P. W. 1979. The role of prostaglandins in platelet function. *Semin. Hematol.* 16:196–207
15. Walsh, P. N. 1974. Platelet coagulant activities and hemostasis: A hypothesis. *Blood* 43:597–605
16. Todd, A. S. 1973. Endothelium and fibrinolysis. *Bibl. Anat.* 12:98–105
17. Moncada, S., Higgs, E. A., Vane, J. R. 1977. Human arterial and venous tissues generate prostacyclin (prostaglandin X), a potent inhibitor of platelet aggregation. *Lancet* 1:18–20
18. Weksler, B. B., Ley, C. W., Jaffe, E. A. 1978. Stimulation of endothelial cell prostacyclin production by thrombin, trypsin, and ionophore A 23187. *J. Clin. Invest.* 62:923–30
19. Ross, R., Glomset, J. A. 1976. The pathogenesis of atherosclerosis. *N. Engl. J. Med.* 295(7):369–77, 420–25
20. Weksler, B. B., Marcus, A. J., Jaffe, E. A. 1977. Synthesis of prostaglandin I_2 (prostacyclin) by cultured human and bovine endothelial cells. *Proc. Natl. Acad. Sci. USA* 74:3922–26
21. Baenziger, N. L., Becherer, P. R., Majerus, P. W. 1979. Characterization of prostacyclin synthesis in cultured human arterial smooth muscle cells, venous endothelial cells and skin fibroblasts. *Cell* 16:967–74
22. Hornstra, G., Haddeman, E., Don, J. A. 1979. Blood platelets do not provide endoperoxides for vascular prostacyclin production. *Nature* 279:66–68
23. Needleman, P., Wyche, A., Raz, A. 1979. Platelet and blood vessel arachidonate metabolism and interactions. *J. Clin. Invest.* 63:345–49
24. Marcus, A. J. 1977. Aspirin and thromboembolism—a possible dilemma. *N. Engl. J. Med.* 297:1284–85
25. Needleman, P. 1979. Prostacyclin in blood vessel—platelet interactions: perspectives and questions. *Nature* 279: 14–15
26. Needleman, P., Kaley, G. 1978. Cardiac and coronary prostaglandin synthesis and function. *N. Engl. J. Med.* 298:1122–28
27. Awbrey, B. J., Hoak, J. C., Owen, W. G. 1979. Binding of human thrombin to cultured human endothelial cells. *J. Biol. Chem.* 254(10):4092–95
28. Czervionke, R. L., Hoak, J. C., Fry, G. L. 1978. Effect of aspirin on thrombin-induced adherence of platelets to cultured cells from the blood vessel wall. *J. Clin. Invest.* 62(4):847–56
29. Hoyer, L. W., de los Santos, R. P., Hoyer, J. R. 1973. Antihemophilic factor antigen: Localization in endothelial cells by immunofluorescent microscopy. *J. Clin. Invest.* 52:2737–44
30. Jaffe, E. A., Hoyer, L. W., Nachman, R. L. 1973. Synthesis of antihemophilic factor antigen by cultured human endothelial cells. *J. Clin. Invest.* 52:2757–64
31. Weiss, H. J., Baumgartner, H. R., Tschopp, T. B., Turitto, V. T., Cohen, D. 1978. Correction by Factor VIII of

the impaired platelet adhesion to subendothelium in von Willebrand disease. *Blood* 51(2):267–79

32. Wall, R. T., Counts, R. B., Harker, L. A., Striker, G. E. 1979. Binding and release of Factor VIII antigen by human endothelial cells. (Submitted)
33. Stemerman, M. B., Ross, R. 1972. Experimental arteriosclerosis. I. Fibrous plaque formation in primates, an electron microscope study. *J. Exp. Med.* 136:769–89
34. Harker, L., Ross, R., Slichter, S., Scott, C. 1976. Homocystine-induced arteriosclerosis: The role of endothelial cell injury and platelet response in its genesis. *J. Clin. Invest.* 58:731–41
35. Minick, C. R., Murphy, G. E. 1973. Experimental induction of athero-arteriosclerosis by the synergy of allergic injury to arteries and lipid rich diet. II. Effect of repeatedly injected foreign protein in rabbits fed a lipid-rich, cholesterol-poor diet. *Am. J. Pathol.* 73:265–300
36. Friedman, R. J., Stemerman, M. B., Wenz, B., Moore, S., Gauldie, J., Gent, M., Tiell, M. L., Spaet, T. H. 1977. The effect of thrombocytopenia on experimental arteriosclerotic lesion formation in rabbits. Smooth muscle cell proliferationan re-endothelialization. *J. Clin. Invest.* 60:1191–1201
37. Fuster, V., Bowie, E. J. W., Lewis, J. C., Fass, D. N., Owen, C. A. Jr. 1978. Resistance to arteriosclerosis in pigs with von Willebrand's disease. Spontaneous and high cholesterol diet-induced arteriosclerosis. *J. Clin. Invest.* 61(3): 722–30
38. Ross, R., Glomset, J., Kariya, B., Harker, L. 1974. A platelet-dependent serum factor that stimulates the proliferation of arterial smooth muscle cells in vitro. *Proc. Natl. Acad. Sci. USA* 71:1207–10
39. Kaplan, K. L., Broekman, J., Chernoff, A., Lesznik, G. R., Drillings, M. 1979. Platelet alpha granule proteins. Studies on release and subcellular localization. *Blood* 53:604–18
40. Antoniades, H. N., Scher, C. D., Stiles, C. D. 1979. Purification of human platelet-derived growth factor. *Proc. Natl. Acad. Sci. USA* 76:1809–13
41. Ross, R., Harker, L. A. 1976. Hyperlipidemia and atherosclerosis: Chronic hyperlipidemia initiates and maintains lesions by endothelial cell desquamation and lipid accumulation. *Science* 193:1094–1100
42. Schwartz, S. M., Haudenschild, C. C., Eddy, E. M. 1978. Endothelial regeneration. I. Quantitative analysis of initial stages of endothelial regeneration in rat aortic intima. *Lab. Invest.* 38:568–80
43. Malczak, H. T., Buck, R. C. 1977. Regeneration of endothelium in rat aorta after local freezing: a scanning electron microscopic study. *Am. J. Pathol.* 86:133–48
44. Fishman, J. A., Ryan, F. B., Karnovsky, M. J. 1975. Endothelial regeneration in the rat carotid artery and the significance of endothelial denudation in the pathogenesis of myointimal thickening. *Lab. Invest.* 30:339–51
45. Stemerman, M. B., Spaet, T. H., Pitlick, F., Cintron, J., Lejnieks, I., Tiell, M. L. 1977. Intimal healing: the pattern of reendothelialization and intimal thickening. *Am. J. Pathol.* 87:125–42
46. Haudenschild, C. C., Schwartz, S. M. 1978. The role of endothelial cell regeneration in the accumulation of intimal smooth muscle cells following trauma. *Fed. Proc.* 37:1387
47. Bell, F. P., Gallus, A. S., Schwartz, C. J. 1974. Focal and regional patterns of uptake and the transmural distribution of ^{131}I-fibrinogen in the pig aorta in vivo. *Exp. Mol. Pathol.* 20:281–92
48. Schwartz, S. M., Benditt, E. P. 1976. Clustering of replicating cells in aortic endothelium. *Proc. Natl. Acad. Sci. USA* 73:651–53
49. Vlodlavsky, I., Fielding, P. E., Fielding, C. S., Gospodarowicz, D. 1978. Role of contact inhibition in the regulation of receptor mediated uptake of low density lipoprotein in cultured vascular endothelial cells. *Proc. Natl. Acad. Sci. USA* 75:356–60
50. Davies, P. F., Ross, R. 1978. Mediation of pinocytosis in cultured arterial smooth muscle and endothelial cells by platelet-derived growth factor. *J. Cell. Biol.* 79:663–71
51. Wall, R. T., Harker, L. A., Striker, G. E. 1978. Human endothelial cell migration. Stimulation by a released platelet factor. *Lab. Invest.* 39:523–29
52. Wall, R. T., Harker, L. A., Quadracci, L. J., Striker, G. E. 1978. Factors influencing endothelial cell proliferation in vitro. *J. Cell. Physiol.* 96:203–14
53. Fass, D. N., Dawning, M. R., Meyers, P., Bowie, E. J., Witte, L. D. 1978. Cell growth stimulation by normal and von Willebrand's porcine platelets and endothelial cells. *Blood* 52(1):181
54. Gajdusek, C., Schwartz, S., DiCorleto, Ross, R. 1979. Endothelial cell derived mitogenic activity. *Fed. Proc.* (Abstr.) 38:1075 (Part II)

55. Wall, R. T., Harlan, J. M., Harker, L. A., Striker, G. E. 1979. In vitro effects of hyperlipoproteinemic plasma on human endothelium: Inhibition of endothelial repair by familial hypercholesterolemic plasma. Submitted for publication
56. Wall, R. T., Harlan, J. M., Harker, L. A., Striker, G. E. 1979. Homocysteine-induced endothelial cell injury in vitro: a model for the study of vascular injury. Submitted for publication
57. Wilcken, D. E. L., Wilcken, B. 1976. The pathogenesis of coronary artery disease: a possible role for methionine metabolism. *J. Clin. Invest.* 57:1079–82
58. Harker, L. A., Slichter, S. J. 1972. Platelet and fibrinogen consumption in man. *N. Engl. J. Med.* 287:999–1005
59. Striker, G. E., Quadracci, L. J., Cutler, R. E., 1978. Use and interpretation of renal biopsy. *Major Problems in Pathology,* Vol. 8, ed. J. L. Bennington. Philadelphia: Saunders
60. Wall, R. T., Harker, L. A., Quadracci, L. J., Striker, G. E. 1977. Immune-mediated endothelial cell injury in the pathogenesis of thrombotic thrombocytopenic purpura (TTP). *Clin. Res.* 25(3):350
61. Amorosi, E. L., Karpatkin, S. 1977. Antiplatelet treatment of thrombotic thrombocytopenic purpura. *Ann. Intern. Med.* 86:102–6
62. Ahmad, S., Young, J. H., Harlan, J., Schribner, B. H., Quadracci, L. J. 1979. Plasma exchange using serum albumin in the management of thrombotic thrombocytopenic purpura. *Clin. Res.* 27(2):287
63. Byrnes, J. J., Lian, E. C. Y. 1979. Recent therapeutic advances in thrombotic thrombocytopenic purpura. *Semin. Thromb. Hemostatis* 5:199–215
64. Kwaan, H. C. 1979. A rational unified approach to the problem of thrombotic thrombocytopenic purpura. *Semin. Thromb. Hemostatis* 5:241–42
65. Fauci, A. S., Haynes, B. F., Katz, P. 1978. The spectrum of vasculitis. Clinical, pathologic immunologic, and therapeutic considerations. *Ann. Intern. Med.* 89:660–76
66. Gaynor, E. 1971. Increased mitotic activity in rabbit endothelium after endotoxin. An autoradiographic study. *Lab. Invest.* 24:318–20
67. Stewart, G. J., Anderson, M. J. 1971. An ultrastructural study of endotoxin induced damage in rabbit mesenteric arteries. *Br. J. Exp. Pathol.* 52:75–80
68. Gerrity, R. G., Richardson, M., Coplan, B. A., Cade, J. F., Hirsh, J., Schwartz, C. J. 1976. Endotoxin-induced vascular endothelial injury and repair. II. Focal injury, en face morphology, ^{3}H-thymidine uptake and circulating endothelial cells in the dog. *Exp. Mol. Pathol.* 24:59–69
69. Raij, L., Keane, W. F., Michael, A. F. 1977. Unilateral Shwartzman reaction; Cortical necrosis in one kidney following in vivo perfusion with endotoxin. *Kidney Int.* 12:91–95
70. Harker, L. A., Slichter, S. J. 1974. Arterial and venous thromboembolism: Kinetic characterization and evaluation of therapy. *Thromb. Diath. Haemorrh.* 31:188–203
71. The Anturane Reinfarction Trial Research Group. 1978. Sulfinpyrazone in the prevention of cardiac death after myocardial infarction. *N. Engl. J. Med.* 298(6):289–95
72. The Canadian Cooperative Study Group. 1978. A randomized trial of aspirin and sulfinpyrazone in threatened stroke. *N. Engl. J. Med.* 299(2):53–59
73. Stewart, G. J., Ritchie, W. G., Lynch, P. R. 1974. Venous endothelial damage produced by massive sticking and emigration of leukocytes. *Am. J. Pathol.* 74:507–32
74. Cooke, E. D., Lloyd, M. J., Bowcock, S. A., Pilcher, M. F. 1977. Intravenous lignocaine in prevention of deep venous thrombosis after elective hip surgery. *Lancet* 2:797–99
75. Harker, L. A., Wall, R. T., Harlan, J. M., Ross, R. 1978. Sulfinpyrazine prevention of homocysteine-induced endothelial cell injury and arteriosclerosis. *Clin. Res.* 26(3):554

Ann. Rev. Med. 1980. 31:373–83
Copyright © 1980 by Annual Reviews Inc. All rights reserved

GIARDIASIS

❖7367

James W. Smith, M.D.

Department of Pathology, Indiana University School of Medicine,
Indianapolis, Indiana 46223

Martin S. Wolfe, M.D.

Office of Medical Services, Department of State, Washington, DC 20520;
Department of Medicine, George Washington Medical School, Washington,
DC 20052; and Department of Medicine, Georgetown Medical School,
Washington, DC 20007

INTRODUCTION

Giardia lamblia was first discovered by Leeuwenhoek (1681), who found the parasite in his own stools. It was long considered a harmless commensal organism but in recent years has been recognized as a cause of intestinal disease often acquired by travelers to foreign countries (1, 2), persons drinking contaminated water in this country (3, 4), children in day care nurseries (5), and homosexual males (6). It is the most common pathogenic intestinal parasite in the United States, being found in 4% of stool specimens submitted to state public health laboratories for parasite examination (7). Attesting to its increasing importance in the United States, a symposium on giardiasis, sponsored by the Environmental Protection Agency, was held in the Fall of 1978 (8).

EPIDEMIOLOGY

In the United States the incidence of *Giardia* varies greatly in different surveys depending on the methods used and the population surveyed. Incidence data from various studies in the US have recently been reviewed (9). The incidence is greater in countries with poor sanitation. The cysts are the important stage in transmission of the infection and were experimentally

0066-4219/80/0401-0373$01.00

shown to survive in water and remain infective for up to 16 days (10). *Giardia* cysts are readily demonstrated in raw sewage but are destroyed by anaerobic digestion in sewage treatment plants (11).

Originally it was thought that there were numerous Giardia species which were quite host specific and that different species infected different mammalian hosts; however, it is now generally agreed that there are two morphologically distinct species which infect mammals (12). One of these, *Giardia duodenalis* (includes *G. lamblia*), naturally infects man, beaver, coyote, cattle, cat, and dog and can experimentally infect some other mammalian species (13). *Giardia muris* infects primarily mice and rats.

Giardiasis is endemic in the United States, but outbreaks are recognized with increasing frequency and certain groups are more likely than others to become infected. Twenty-three waterborne outbreaks of giardiasis have been reported in the United States since 1965 (3). Campers have been infected by drinking untreated mountain stream water in Utah (14) and Colorado (15) and beavers have been implicated as a reservoir (4, 13). The beavers have probably acquired the infection from man and have served as an amplification reservoir.

Several large outbreaks have occurred as a result of raw sewage contaminating municipal water supplies [Aspen, Colorado (16) and Vail, Colorado (17)] or from inadequate water treatment systems [Rome, New York (18) and Washington, New Hampshire (19)]. Chlorination alone, while effective in killing enteric bacteria, may not always kill *Giardia* cysts. Proper sedimentation and filtration are required to remove cysts from surface water. In one of the outbreaks (18) cysts were demonstrated in treated water and shown to be infective for dogs.

Giardiasis, probably acquired from water, is a common problem in travelers returning from foreign countries. Increased risk for contracting infection has been shown for the Soviet Union (particularly Leningrad) (1), Mexico, Western South America, Southeast Asia, the Middle East, the Far East, and Southern Asia (2).

Giardiasis has also been a problem in day care centers where it is probably spread by fecal-oral contamination (5). Incidence is high in homosexual males in whom it is associated with the practice of anilingus (6). Infection can also be spread by contaminated food (20).

PATHOGENESIS

Epidemiologic and experimental data suggest that individuals and various strains of experimental animals (21) differ in their susceptibility to infection, and that strains of *G. lamblia* differ in virulence. Immunocompromised individuals are particularly likely to develop serious long-lasting infection

(22). Experimental studies have shown that as few as 10 cysts can infect man (23) and one cyst can infect the mouse (24). The organisms excyst in the stomach or upper small intestine. Persons with achlorhydria appear to be more susceptible to infection (25). Trophozoites infect the duodenum and upper jejunum and are most numerous at the bases of the microvilli where they live in the unstirred mucus layer or attached to the mucosa by the ventral disc (26). Attachment is achieved by contraction of the periphery of the disc (27) and/or hydrodynamic forces produced by the ventral flagella (28). Factors that lead to release of attached organisms and to encystment are not clear. Patients may have numerous trophozoites in the upper intestine without cysts in feces, particularly early in infection (29). Some patients regularly pass large numbers of cysts in stools, some pass low numbers, and some pass variable numbers in a cyclic fashion (30). Infection is limited to the upper small intestine but the reasons are not clear. One suggestion is that the alkaline pH may be more favorable (31). Nutritional factors or specific receptors on the brush border of mucosal cells (32) have also been proposed.

Several theories for the pathogenicity of *Giardia* have been proposed. Mechanical blockage of the brush border by numerous parasites was suggested (33), but much of the brush border is not covered by organisms (26), the majority being unattached in the mucus layer. Mechanical irritation of mucosa by organisms may be important but has not been proven. Production of toxic substances by the organisms has been suggested; however, culture filtrates do not cause intestinal damage in the rabbit (34). Bile salt abnormalities have been proposed (35) but not confirmed (34). Associated bacterial overgrowth may contribute to severe malabsorption in some but not all patients (36). Treatment to eradicate the bacteria without treating the *Giardia* has not been effective (37). A recent study from India has suggested a possible interrelation between *G. lamblia* and jejunal colonization with *Candida albicans* (38).

The role of the immune system in the pathogenesis of giardiasis has been studied in man and in mouse models and recently summarized (39). Both humoral and T-cell immunity appear to be important. Penetration of *Giardia* trophozoites through breaks in the mucosa (32) or by actual invasion (40) to intestinal lymphoid tissue may be important in the development of an immune response. Mice infected with large inocula of *G. muris* had a more rapid onset of infection, were sicker, and cleared the infection more rapidly than those receiving small inocula (41). Those with the shortest primary infections had the greatest resistance to reinfection. Increased resistance to reinfection has also been noted in man but reinfection may occur. Infection in nude (athymic) mice persisted for a much longer time than in immunologically intact mice (42). Nude mice reconstituted with lymphoid cells from nonimmune mice cleared the infection faster than nude controls. Those reconstituted from previously infected mice cleared it even faster (21). Thymus-derived lymphocytes have been noted to cross the epithelium and attach to trophozoites in the lumen.

These observations suggest that thymus-derived lymphocytes are important in clearing the infection. After treatment, nude mice were susceptible to reinfection although the infecting dose required was higher than for nude controls that had never been infected; this suggests that humoral immunity also plays a role (42). Antibody can be demonstrated in some persons infected with *Giardia* (43, 44). The exact role of the various immunoglobulins is not clear (39).

Pathologic findings vary greatly (26, 34) ranging in biopsies from normal intestinal morphology to changes resembling sprue. Abnormalities may include decreased villus-crypt ratio, increased epithelial cell turnover, local acute inflammation, round cell infiltration, and hyperplasia of lymphoid nodules. Intercellular inflammatory cells may be present in the mucosa. The severity of morphologic changes correlates with the clinical severity of the disease (34) in some but not all patients. Although there was some morphologic change in the mucosa of infected nude mice, if the mice were immunologically reconstituted with lymphoid cells from previously infected mice, the morphologic changes were much more severe with a decreased villus-crypt ratio and increased cellular infiltration. This shows that the T-cell system is important in the pathogenesis of mucosal lesions (21). Much still remains to be learned about the pathogenesis of this infection.

CLINICAL DISEASE

The incubation period varies but is generally one to three weeks and averages 9.1 days (10). *Giardia* infection can be asymptomatic but may cause various clinical manifestations (45–47) ranging from mild transient intestinal complaints that resolve spontaneously to severe long-standing disease with malabsorption and debilitation.

Acute infection is most severe and is characterized by the sudden onset of explosive, foul-smelling, watery diarrhea that is free of blood or pus. There is abundant intestinal gas manifested by abdominal distention, foul-smelling flatulence, and/or belching. Abdominal cramps are common and usually are upper to mid epigastric (in contrast to most other causes of diarrhea, which cause periumbilical or lower abdominal cramps). Greasy, frothy stools that float on toilet water are typical. Anorexia and nausea are common and vomiting occurs in some patients. Systemic manifestations may include generalized weakness, chills, and low grade fever. Weight loss is common. The duration of acute clinical manifestations varies greatly. In some patients they last for only three or four days and in others for months. In most patients the acute disease lasts one to four weeks. The acute infection commonly resolves spontaneously and organisms disappear from the feces. Some patients become carriers for a period and have no additional clinical manifestations. Still others may have recurrences of clinical symptoms. In occasional patients acute infection persists and leads to malabsorption syndrome with steatorrhea, lassitude, and marked weight loss. Some patients may have a persistent lactase deficiency after treatment and symptoms of this should not be confused with recurrence of infection.

Subacute infection is characterized by mild or moderate symptoms, often without history of more severe acute infection, and may last for months or years. It is characterized by intermittent episodes of mushy foul-smelling stools with mid epigastric pain or substernal burning, and increased distention, foul-smelling flatulence, and belching. There are often anorexia, weight loss, fatigue, and a general unwell feeling.

Chronic infection is a mild, persistent infection that may or may not have been preceded by an acute or subacute infection. It is characterized by periodic episodes of mushy foul-smelling stools with increased flatulence and abdominal distention. There may be substernal burning but cramps are uncommon. Anorexia is sometimes noted during exacerbations.

Subacute and chronic infection are likely to be confused with peptic ulcer, gallbladder disease, or hiatal hernia; giardiasis should be considered in the differential diagnosis of these conditions.

DIAGNOSIS

The principal method of diagnosis is demonstration of the organism in fecal specimens or material from the upper small intestine. Fecal examination is the easiest and will allow diagnosis of most infections. However, stools may be negative in the early stages of infection (29) and in patients who shed organisms in a cyclic pattern (30). Three separate stools, one collected every second or third day, are recommended. Purgation does not offer any advantages. Three stool examinations will detect over 90% of infections that can be diagnosed by stool examination. Stools generally contain only cysts but the more fragile trophozoites are sometimes found, especially in diarrheic specimens.

Formed specimens, which contain the more hardy cysts, can be examined within 24 hours (longer if refrigerated) or can be placed in fixatives. However, diarrheic specimens should either be examined within one hour or should be placed in fixatives such as polyvinyl alcohol fixative (for permanent stains) and formalin or merthiolate iodine formalin (for direct examination and concentration).

Most reported studies have examined fresh or fixed specimens by only direct examination and concentration. However, a recent report (48) found that of 21 positive specimens, organisms were demonstrated in five by chlorazol black and in three by trichrome permanent stains but not by direct examination or concentration methods. This study suggests that use of permanent stains of feces may improve the detection of *Giardia.* Barium, antibiotics, antacids, kaolin products, paragoric, many enema preparations and oily laxatives may interfere with laboratory procedures or may lead to disappearance of parasites from stool. Examination of the stool or upper intestinal material should be delayed for 5 to 10 days after the patient has received these substances.

Patients suspected of having giardiasis who have negative stool examination may have trophozoites in the upper small intestine which can be demonstrated in intestinal fluid or biopsy. Upper intestinal fluid can be obtained by passing a tube into the small intestine and aspirating, or by a commercially available string test (49) (Enterotest® HEDECO Company, Mountain View, California), in which a string tied to a weighted capsule is swallowed with the proximal end taped to the patient's cheek. The string is later retrieved and fluid from the bile-stained distal portion is examined. Intestinal fluid may be examined both by wet mounts and permanent stains. It is best to contact the laboratory prior to submitting the specimen and to obtain the specimen during regular working hours. These specimens must be transported to the laboratory promptly. A recent publication suggests that secretin stimulation prior to obtaining intestinal fluid specimens may increase the yield of positive material (50). Upper small intestinal biopsy may be examined by preparing Giemsa-stained smears of the mucosal surface and/or by preparing histologic sections and examining them for organisms and pathologic changes (51).

It is important that the laboratory personnel be competent in identification of intestinal protozoa. Results of a survey program show that significant numbers of laboratories have difficulty identifying *Giardia* cysts in wet mounts (52); however, there has been improvement in recent years.

Occasional patients have a strong clinical and epidemiologic basis for suspecting giardiasis, but organisms cannot be demonstrated by any of the above methods. Empiric treatment with quinacrine hydrochloride should be considered before giardiasis is ruled out (45).

Serologic methods have recently been reported (43, 44). An indirect fluorescent antibody test using axenically cultured *G. lamblia* (44) is being developed by the Center for Disease Control and shows promise. Hopefully, a useful serologic test will be available before too long.

Although radiographic abnormalities of an inflammatory bowel disease pattern are seen in some patients, in most the small bowel appears normal.

Patients with malabsorption syndrome may have radiographic changes similar to those in sprue (53).

When specifically tested, patients often have laboratory evidence of malabsorption (37) with lactase deficiency, steatorrhea, low serum carotene, abnormal D-xylose excretion, and abnormal B_{12} absorption. However, generally the above studies are not indicated as they improve following specific treatment.

TREATMENT

Quinacrine hydrochloride (Atabrine®) is considered the treatment of choice for giardiasis by most workers. The recommended dose for those 8 years old and above is 100 mg three times a day for 7 days. For younger children, the recommended dose is 2 mg/kg three times a day for 7 days. Cure rates of at least 90% are reported by most workers (45, 54). One remarkable exception to these excellent results is a recent report from London where *Giardia* was eradicated in only 5 of 8 (63%) patients given this drug (37). Side effects are not uncommon with quinacrine; these may include nausea, vomiting, headaches, dizziness, and yellow discoloration of the urine (45, 54). Rarely, fever and severe skin rash may occur. A reversible toxic psychosis of either depression or excitation may occur, particularly in adults (2), and patients should be forewarned of this possibility. Yellowing of the skin and sclerae rarely occur in doses used for giardiasis. Alcohol is contraindicated with quinacrine use, and quinacrine should not be taken by psoriatics, or taken together with primaquine.

Metronidazole (Flagyl®) is frequently prescribed for giardiasis. In this country the usual recommended adult dose is 250 mg three times a day for 7 days, and for children 5 mg/kg three times a day for 7 days. Cure rates in these doses in adults are somewhat less than with quinacrine, in the range of 86% (45). It appears that metronidazole in the usual recommended dose is less effective for acute than established infections (45, 55). This is perhaps related to the excellent absorption of this drug. In England, metronidazole in a single 2.0-g dose for 3 successive days produced a parasitological cure rate of 91% (37). As metronidazole is unapproved by the Food and Drug Administration for treatment of giardiasis in the United States, these higher doses cannot be recommended; also, the effectiveness and tolerance of this regimen are no better than with quinacrine. A recent comparison of four dosage schedules carried out in Finland on returnees from the Soviet Union showed the highest cure rate (93%) to be with a course of 200 mg three times a day for 7 days followed by an identical course given after a 7 day interval (56). Tolerance to metronidazole is generally better than with quinacrine. Side effects may include metallic taste, dark discoloration of urine, nausea, and headache (45, 54). Monilial infections of the vagina or gastrointestinal tract may follow after metronidazole treatment. Metronidazole has a well-recognized disulfuram-like reaction when taken with alcohol. Metronidazole has been found to be carcinogenic in mice, but not in hamsters; it is mutagenic in bacteria and there is controversy concerning its use in non-life-threatening infections (57).

Furazolidone (Furoxone®) has been found to be effective and well tolerated in studies abroad (58, 59). It is the only anti-*Giardia* drug available as a suspension in the United States, making it particularly useful in young children. Side effects, however, are not uncommon. These include vomiting, diarrhea, nausea, and fever. Hypersensitivity reactions have occurred rarely, such as hypotension, urticaria, and serum sickness. Furazolidone may cause mild reversible hemolysis in glucose-6-phosphate dehydrogenase-deficient individuals. It is a monoamine oxidase inhibitor and can precipitate a disulfuram-like reaction with alcohol. Brown discoloration of the urine may occur, caused by a metabolite. Also troublesome has been the induction of mammary tumors in rats, which led the Food and Drug Administration to question its use in humans; with satisfactory warnings included in a revised package insert, the drug is approved in the United States for treatment of giardiasis. Recommended dosage for adults is 100 mg four times a day and for children 6 mg/kg/day in four divided doses, for 7 days.

A number of other drugs including tinidazole (Fasigyn®) (55) and nimorazole (59) have been found useful abroad, but these drugs are not licensed or available in the United States.

None of the three available drugs in this country has proved safe in pregnancy, and if used at all during pregnancy they should be administered only when severe symptoms are definitely attributable to giardiasis and benefit is judged to outweigh potential risk. Quinacrine has been used in a small number of highly symptomatic pregnant women without recognizable deleterious effects on the fetus (47).

There is some difference of opinion over the necessity of treating asymptomatic cyst-passers. Since the potential exists for the development of intermittent chronic symptoms in asymptomatic cases, and as there is a risk of asymptomatic foodhandlers and children infecting others (5), a strong case can be made for treating all those infected who have no contraindication, such as pregnancy, to the use of the available drugs.

PREVENTION

Chemoprophylactic drugs are neither safe nor effective for giardiasis. Infection has occurred even in those who apparently take all recommended precautions regarding contaminated water or food. Nevertheless, in endemic areas water for drinking or for ice should be boiled for 10 minutes or purified with iodine compounds (60). Uncooked or unpeeled fruits and vegetables should be avoided. Though prepared food has not been so definitively implicated as water as a source of *Giardia* infection, it is likely that cold foods and salads contaminated with infected water or fingers must lead to infection, and reliance should be placed on hot cooked foods. A study of travelers to Madeira Island implicated ice cream and raw vegetables as sources of infection (20). In some waterborne giardiasis outbreaks, water has been found to be coliform-free and it is possible that though chlorination may be sufficient to destroy these indicator organisms, *Giardia* cysts may require long contact time for their inactivation (4). Data at this time indicate that well-operated conventional treatment plants employing coagulation/flocculation, settling, and filtration are necessary to prevent waterborne outbreaks (4).

Literature Cited

1. Walzer, P. D., Wolfe, M. S., Schultz, M. G. 1971. Giardiasis in travelers. *J. Infect. Dis.* 124:235–37
2. Wolfe, M. S. 1978. Current concepts in parasitology. Giardiasis. *N. Engl. J. Med.* 298:319–21
3. Craun, G. F. 1979. Waterborne outbreaks of giardiasis. See Ref. 8
4. Center for Disease Control. 1977. Foodborne and waterborne disease outbreaks, 1976 Annual Summary. Atlanta, Ga.
5. Black, R. E., Dykes, A. C., Sinclair, S. P., Wells, J. G. 1977. Giardiasis in day-care centers: evidence of person to person transmission. *Pediatrics* 60:486–91
6. Schmerin, M. J., Jones, T. C., Klein, H. 1978. Giardiasis: association with homosexuality. *Ann. Intern. Med.* 88: 801–3
7. Center for Disease Control. 1978. *Intestinal Parasite Surveillance, Annual Summary 1977.* Issued September 1978. Atlanta, Ga.
8. Jacobowski, W., Hoff, J. C., eds. 1979. *Proc. Natl. Symp. on Waterborne Transmission of Giardiasis.* US Environ. Protect. Agency, Cincinnati, Ohio. Sept. 18–20, 1978. (EPA-600/9-79-001) Washington, DC: GPO. 306 pp. Available through NTIS, Springfield, Va. 22151
9. Healy, G. R. 1979. The presence and absence of *Giardia lamblia* in studies on parasite prevalence in the USA. See Ref. 8
10. Rendtorff, R. C., Holt, C. J. 1954. The experimental transmission of human intestinal protozoan parasites. IV. Attempt to transmit *Endamoeba coli* and *Giardia lamblia* cysts by water. *Am. J. Hyg.* 60:327–38
11. Fox, J. C., Fitzgerald, P. R. 1979. The presence of *Giardia lamblia* cysts in sewage and sewage sludges from the Chicago area. See Ref. 8
12. Levine, N. D. 1973. *Protozoan Parasites of Domestic Animals and of Man,* pp. 118–22. Minneapolis, Minn: Burgess. 406 pp.
13. Davies, R. B., Hibler, C. P. 1979. Animal reservoirs and cross-species transmission of *Giardia.* See Ref. 8
14. Barbour, A. G., Nichols, C. R., Fukushima, T. 1976. An outbreak of giardiasis in a group of campers. *Am. J. Trop. Med. Hyg.* 25:384–89
15. Wright, R. A., Spencer, H. C., Brodsky, R. E., Vernon, T. M. 1977. Giardiasis in Colorado: An epidemiologic study. *Am. J. Epidemiol.* 105:330–36
16. Moore, G. T., Cross, W. M., McGuire, D., Mollahan, C. S., Gleason, N. N., Healy, G. R., Newton, L. H. 1969. Epidemic giardiasis at a ski resort. *N. Engl. J. Med.* 281:402–7
17. Center for Disease Control. 1978. Giardiasis—Vail, Colorado. Morbidity and Mortality Weekly Report. 27:155
18. Shaw, P. K., Brodsky, R. E., Lyman, D. O., Wood, B. T., Hibler, C. P., Healy, G. R., Macleod, K. I. E., Stahl, W., Schultz, M. G. 1977. A communitywide

outbreak of giardiasis with evidence of transmission by a municipal water supply. *Ann. Intern. Med.* 87:426–32

19. Center for Disease Control. 1977. Waterborne giardiasis outbreaks—Washington, New Hampshire. Morbidity and Mortality Weekly Report. 27:169–170, 175
20. Lopez, C. E., Juranek, D. D., Sinclair, S. P., Schultz, M. G. 1978. Giardiasis in American travelers to Madiera Island, Portugal. *Am. J. Trop. Med. Hyg.* 27:1128–32
21. Roberts-Thomson, I. C., Mitchell, G. F. 1978. Giardiasis in mice I. Prolonged infections in certain mouse strains and hypothymic (nude) mice. *Gastroenterology* 75:42–46
22. Ament, M. E., Rubin, C. E. 1972. Relation of giardiasis to abnormal intestinal structure and function in gastrointestinal immuno-deficiency syndrome. *Gastroenterology* 62:216–26
23. Rendtorff, R. C. 1954. Experimental transmission of human intestinal parasites. II. *Giardia lamblia* cysts given in capsules. *Am. J. Hyg.* 59:209–20
24. de Carneri, I., Trane, F., Mandelli, V. 1977. *Giardia muris:* oral infection with one trophozoite and generation time in mice. *Trans. R. Soc. Trop. Med. Hyg.* 71:438
25. Gianella, R. A., Broitman, S. A., Zamcheck, N. 1973. Influence of gastric acidity on bacterial and parasitic enteric infections. *Ann. Intern. Med.* 78:271–76
26. Yardley, J. H., Takano, J., Hendrix, T. R. 1964. Epithelial and other mucosal lesions of the jejunum in giardiasis—Jejunal biopsy studies. *Bull. Johns Hopkins Hosp.* 115:389–406
27. Erlandsen, S. L., Feeley, D. E., Schollmeyer, J. V., Chase, D. G. 1979. In vitro model for analyzing the attachment/release of trophozoites of *Giardia:* evidence for the involvement of contractile proteins and the effect of atabrine. See Ref. 8
28. Holberton, D. V. 1974. Attachment of *Giardia:* a hydrodynamic model based on flagellar activity. *J. Exp. Biol.* 60:207–21
29. Jokipii, A. M. M., Jokipii, L. 1977. Prepatency of giardiasis. *Lancet* 1 (8021):1095–97
30. Danciger, M., Lopez, M. 1975. Numbers of *Giardia* in the feces of infected children. *Am. J. Trop. Med. Hyg.* 24:237–42
31. Haiba, M. H. 1954. The pH of the alimentary tract in normal and giardia infected culture mice. *Parasitology* 44: 387–91
32. Owen, R. L., Nemanic, P. C., Stevens, D. P. 1979. Ultrastructural observations on giardiasis in a muris model. I. Intestinal distribution, attachment and relationship to the immune system in *Giardia muris. Gastroenterology* 76:757
33. Veghelyi, P. 1939. Coeliac disease initiated by giardiasis. *Am. J. Dis. Child.* 57:894–99
34. Duncombe, V. M., Bolin, T. D., Davis, A. E., Cummins, A. G., Crouch, R. L. 1978. Histopathology in giardiasis—a correlation with diarrhoea. *Aust. N. Zealand J. Med.* 8:392–96
35. Tandon, B. N., Tandon, R. K., Satpathy, B. K., Shriniwas. 1977. Mechanism of malabsorption in giardiasis: a study of bacterial flora and bile salt deconjugation in upper jejunum. *Gut* 18:176–81
36. Tomkins, A. M., Wright, S. G., Drasar, B. S., James, W. P. T. 1978. Bacterial colonization of jejunal mucosa in giardiasis. *Trans. R. Soc. Trop. Med. Hyg.* 72:33–36
37. Wright, S. G., Tomkins, A. M., Ridley, D. S. 1977. Giardiasis: clinical and therapeutic aspects. *Gut* 18:343–50
38. Naik, S. R., Rau, N. R., Vinayak, V. K., Narayanan, V. A., Zungurwade, S., Sehgal, S. C., Talwar, P. 1978. Presence of *Candida albicans* in normal and in *Giardia lamblia* infected human jejunum. *Ann. Trop. Med. Parasitol.* 72: 493–94
39. Kraft, S. C. 1979. The intestinal immune response in giardiasis (Editorial). *Gastroenterology* 76:877
40. Brandborg, L. L., Tankersley, C. B., Gottlieb, S., Barancik, M., Sartor, V. E. 1967. Histological demonstration of mucosal invasion by *Giardia lamblia* in man. *Gastroenterology* 52:143–50
41. Roberts-Thomson, I. C., Stevens, D. P., Mahmoud, A. A. F., Warren, K. S. 1976. Giardiasis in the mouse: an animal model. *Gastroenterology* 71:57–61
42. Stevens, D. P., Frank, D. M., Mahmoud, A. A. F. 1978. Thymus dependency of host resistance to *Giardia muris* infection: studies in nude mice. *J. Immunol.* 120:680–82
43. Ridley, M. J., Ridley, D. S. 1976. Serum antibodies and jejunal histology in giardiasis associated with malabsorption. *J. Clin. Pathol.* 29:30–34
44. Visvesvara, G. S., Healy, G. R. 1979. The possible use of an indirect immunofluorescent test using axenically grown *Giardia lamblia* antigens in diagnosing giardiasis. See Ref. 8

45. Wolfe, M. S. 1975. Giardiasis. *J. Am. Med. Assoc.* 233:1362–65
46. Yardley, J. H., Bayless, T. M. 1967. Giardiasis. *Gastroenterology* 52:301–4
47. Wolfe, M. S. 1979. Managing the patient with giardiasis: clinical, diagnostic, and therapeutic aspects. See Ref. 8
48. Gleason, N. N., Horowitz, M. S., Newton, L. H., Moore, G. T. 1971. A stool survey for enteric organisms in Aspen, Colorado. *Am. J. Trop. Med. Hyg.* 19:480–84
49. Beal, C. B., Viens, P., Gart, R. G. L., Hughes, J. M. 1970. A new technique for sampling duodenal contents. *Am. J. Trop. Med. Hyg.* 19:349–52
50. Sagaro, E., Blanco, E., Fragoso, T., Castaneda, C. 1977. Duodenal intubation with secretin stimulus for diagnosis of giardiasis. *Arch. Dis. Child.* 52:505–7
51. Kamath, K. R., Murugasu, R. 1974. A comparative study of four methods for detecting *Giardia lamblia* in children with diarrheal disease and malabsorption. *Gastroenterology* 66:16–21
52. Smith, J. W. 1979. Morphologic diagnosis of fecal parasitic infections in the United States. *Am. J. Clin. Pathol.* 72:371–73.
53. Fisher, C. H., Oh, K. S., Bayless, T. M., Siegelman, S. S. 1975. Current perspectives on giardiasis. *Am. J. Roentgenol. Radium Ther. Nucl. Med.* 125:207–17
54. Kavousi, S. 1979. Giardiasis in infancy and childhood: a prospective study of 160 cases with comparison of quinacrine (Atabrine®) and metronidaxole (Flagyl®) *Am. J. Trop. Med. Hyg.* 28: 19–23
55. Andersson, T., Forssell, J., Sterner, G. 1972. Outbreak of giardiasis: effect of a new anti-flagellate drug, tinidazole. *Br. Med. J.* 2:449–51
56. Jokipii, L., Jokipii, A. M. M. 1978. Comparison of four dosage schedules in the treatment of giardiasis with metronidazole. *Infection* 6:92–94
57. Is Flagyl dangerous? 1975. *Med. Lett. Drugs Ther.* 17:53–54
58. Bassily, S., Farid, A., Mikhail, J. W., Kent, D. C., Lehman, J. S. 1970. The treatment of *Giardia lamblia* infection with mepacrine, metronidazole, and furazolidone. *J. Trop. Med. Hyg.* 73: 15–18
59. Levi, G. C., de Avila, C. A., Neto, V. A. 1977. Efficacy of various drugs for treatment of giardiasis. A comparative study. *Am. J. Trop. Med. Hyg.* 26: 564–65
60. Kahn, F. H., Visscher, B. R. 1975. Water disinfection in the wilderness. A simple, effective method of iodination. *West. J. Med.* 122:450–53

Ann. Rev. Med. 1980. 31:385–404
Copyright © 1980 by Annual Reviews Inc. All rights reserved

TREATMENT OF HYPERTHYROIDISM

❖7368

Michael R. McClung, M.D. and Monte A. Greer, M.D.

Division of Endocrinology, Department of Medicine, University of Oregon
Health Sciences Center, Portland, Oregon 97201

INTRODUCTION

During the past decade, our understanding of thyroid pathophysiology has expanded greatly. Recent advances include insights into the mechanisms of thyroid hormone action, the peripheral metabolism of thyroid hormones, and immunologic aspects of thyroid disease. Refinements in thyroid hormone assays and the availability of thyrotropin-releasing hormone (TRH) have provided more direct and definitive approaches to the diagnosis of hyperthyroidism. Using these new laboratory aids, researchers recently demonstrated that hyperthyroidism is more common than was formerly believed, occurring in as much as 2% of the adult female population (1, 2). However, the therapeutic tools with which we approach hyperthyroid patients have changed little in the past 30 years. Each of the three principal methods, antithyroid drugs, radioiodine, and surgery, can be used effectively and each has its advantages, its advocates, and its tradition in the therapy of hyperthyroid individuals. The management of a specific patient will be influenced by the nature and severity of the hyperthyroidism, the patient's age and general health, the experience of the responsible physician, the availability of expert nuclear medicine and surgical colleagues, and by the patient's own feelings regarding therapy.

THERAPEUTIC APPROACHES

Since the etiologies of the various types of thyroid hormone overproduction are not yet known, none of the available treatments is directed specifically at a basic pathophysiologic defect. Therapeutic options available include

0066-4219/80/0401-0385$01.00

(*a*) inhibiting thyroid hormone synthesis, (*b*) inhibiting thyroid secretion, (*c*) reducing the mass of functioning thyroid tissue, (*d*) decreasing the peripheral effects of thyroid hormone.

Decreased Thyroid Hormone Synthesis and/or Secretion

ANTITHYROID DRUGS The antithyroid drugs commonly in use throughout the world are generically thionamides. All have an $=\mathrm{N}-\underset{|}{\mathrm{C}}=\mathrm{S}$ configuration in the molecule. The two most commonly employed substances in the United States are propylthiouracil and methimazole, which have 6- and 5-membered ring structures, respectively. It is believed that thionamides act primarily by inhibiting thyroid peroxidase, an enzyme that catalyzes both the oxidation of iodide ion to a form in which it can be covalently bound into the tyrosyl residues in thyroglobulin, and the "coupling" of iodotyrosyl residues in thyroglobulin to form thyroxine and triiodothyronine (3). The coupling reaction is much more sensitive to inhibition by thionamides (4, 5). Thus a complete remission of thyrotoxicosis can be induced by antithyroid drugs, even though a supranormal thyroid radioiodine uptake is maintained. Thionamide drugs inhibit the iodination of thyroglobulin and the synthesis of thyroid hormone but do not inhibit the thyroid iodide pump, which concentrates iodide in the thyroid cell, nor do they directly affect the rate of secretion of preformed thyroid hormone stored within the thyroid follicles (3).

There is a large store of hormone within the thyroid gland, enough in a normal individual with a normal rate of secretion to maintain a euthyroid state for 6–9 months even if all synthesis of thyroid hormone is abolished. Therefore an immediate return to a euthyroid state cannot be achieved in thyrotoxic patients treated solely with these drugs. Although patients with thyrotoxicosis usually have enlarged thyroids and a supranormal amount of stored hormone, their rate of secretion of thyroid hormone is greatly increased above normal. Thus, their plasma thyroid hormone concentration falls to a euthyroid range in an average of 3 months, a considerably more rapid rate of decline than can be achieved in normal individuals.

The relative potency of propylthiouracil and methimazole is not definitively established. Early data, employing acute suppression of radioiodine uptake as an index, suggested that methimazole was 100 times as potent as propylthiouracil (6). Subsequent clinical trials indicated that the potency ratio might be 10:1. The standard dose is 300 mg propylthiouracil or 30 mg methimazole daily. It is usually recommended that the drugs be given in 3 equal doses at 8-hour intervals because of their rapid metabolism. This is inconvenient and compliance with an 8-hour schedule is difficult to achieve. In spite of theoretical objections, we have found that in almost all patients, including those severely toxic, giving either propylthiouracil or methimazole once daily in the same total daily dose that would be used if the medication were given every 8 hours, induces a remission of the thyrotoxicosis as rapidly and effectively as giving divided doses (7, 8). We have routinely used the single-daily-dose regimen in our clinic for the last 20 years. Only rarely have patients failed to respond to this program. In such resistant patients, it is not clear that divided doses are more effective, since they are usually also refractory to divided daily doses. We know of only one report indicating that divided daily doses are more effective (9). Since there is considerable variation among patients in the rapidity of their response to antithyroid drugs, it is possible that patients who do not become euthyroid until they are changed from single to divided daily doses might have had the same temporal response to therapy if the regimen had not been changed.

In spite of its simplicity and demonstrated efficacy, once-a-day administration of antithyroid drugs is not widely employed. This is probably because physicians are reluctant to change from a therapeutic plan they know will work to a potentially improved technique that would be experimental for them. Although we have not found it necessary, the physician may feel more comfortable in prescribing antithyroid drugs on an every-8-hours rather than a single-daily-dose regimen for severely thyrotoxic patients. The dose can be given once daily as soon as the patient shows definite improvement.

Serious complications from antithyroid drugs are quite rare, agranulocytosis being the most dangerous (10). Routine weekly blood counts are of little prophylactic value, since agranulocytosis may develop within a few hours (11). It is important to warn the patient to discontinue the drug and to contact the physician if fever, malaise, or sore throat develops. There is usually spontaneous recovery of leukocyte production within 1–2 weeks after stopping the drug. With appropriate antimicrobial and supportive therapy, agranulocytosis is rarely fatal. Radioiodine therapy is employed once bone marrow function is restored.

Minor drug toxicity, such as skin rash, pruritis, or arthralgia, occurs in 3–5% of patients treated with thionamide drugs (12). This may subside spontaneously if the drug is continued, or it may disappear if one switches to the other drug (e.g. from propylthiouracil to methimazole). If the problem persists, antithyroid drug therapy is discontinued and an alternative form of therapy chosen.

IODIDE The effect of large doses of iodide to decrease thyroid secretion and vascularity has been employed since its introduction by Plummer for the preoperative management of hyperthyroid patients. Except in this situation, iodides should not be used alone in thyrotoxicosis. Large doses of iodide can result in increased synthesis of thyroid hormone because of a marked increase in the available substrate (one drop of Lugol's solution contains 6 mg of iodine, compared to a normal daily intake of approximately 200 μg). Iodide does not induce a decrease in thyroid secretion in all thyrotoxic patients. Radioiodine therapy is not possible while a patient is being given iodide because the marked increase in the body iodide pool prevents effective concentration of radioiodine in the thyroid. By increasing intraglandular stores of thyroid hormone, iodide administration increases the time required to see the effect of subsequent thionamide therapy. However, the administration of stable iodide following the initiation of adequate antithyroid or radioiodine therapy may more rapidly improve both chemical and clinical parameters of hyperthyroidism (13, 14) and is usually effective in patients with severe thyrotoxicosis. Iodide therapy is rarely employed since the advent of propranolol a much more useful drug.

LITHIUM The antithyroid effect of lithium became apparent when goiter and/or hypothyroidism developed in about 4% of the patients treated with lithium carbonate for affective disorders (15). The predominant effect on the thyroid, like that of iodide, is to decrease secretion of the preformed hormones, although very large doses may also decrease thyroid hormone synthesis (16). The antithyroid effect of lithium is more pronounced in hyperthyroid than in euthyroid patients, and lithium alone has been used successfully to treat thyrotoxicosis (17, 18).

Effective serum concentrations of lithium are 0.5–1.2 meq/liter. Toxicity is often noted at concentrations greater than 1.5 meq/liter. Because of this very narrow therapeutic ratio, lithium therapy in hyperthyroidism requires careful monitoring of serum lithium levels and of the patient's clinical status. In a controlled study comparing methimazole and lithium carbonate, Kristensen and co-workers described side effects in 8 of 11 lithium-treated patients while noting that lithium offered no significant benefit over methimazole in the early clinical or chemical response to therapy (19).

In patients for whom very rapid improvement of hyperthyroidism is necessary, a combination of lithium and thionamides has been employed (20). This approach is as effective as the combined use of iodide and antithyroid drugs and offers the advantage of not delaying the subsequent use of radioiodine therapy. Although we do not advocate the routine use of lithium in the treatment of hyperthyroidism, it may be a very useful agent in a patient who is allergic to antithyroid drugs and who is awaiting radioiodine therapy, or in conjunction with thionamide therapy in some patients with severe hyperthyroidism for whom rapid control is important.

Thyroid Ablation

RADIOACTIVE IODINE Radioactive iodine provides a simple, effective, inexpensive, and safe treatment for hyperthyroidism. By virtue of its unique capacity to transport and to bind iodine in the form of iodothyronines, the thyroid gland can be selectively irradiated. ^{131}I is the only isotope of iodine currently in wide use for the treatment of hyperthyroidism. The beta emissions of this isotope are responsible for the majority of the thyroidal radiation. Although theoretical arguments suggested that ^{125}I would result in a lower incidence of hypothyroidism, this was not the case in practice, and the original proponents of ^{125}I therapy have now abandoned its use (21).

Radioiodine concentration by the thyroid must occur if ^{131}I therapy is to be effective, and demonstration of adequate thyroid uptake prior to therapy is necessary. Radioiodine administration is contraindicated during pregnancy in order to avoid fetal injury.

Concern is often expressed over the possibility of inducing malignancy. Isolated cases of thyroid cancer, hematologic malignancies, and other neoplasms have occurred in patients previously treated with ^{131}I for hyperthyroidism. However, a large group of patients followed up to 25 years after radioiodine therapy has no greater incidence of leukemia or thyroid cancer than a surgically treated group (22).

The radiation dose to the gonads in patients treated with ^{131}I is usually less than 3 rads. This is not significantly different from gonadal irradiation received with several commonly used radiographic diagnostic tests such as barium enemas or pyelography. We agree with the conclusion of Robertson & Gorman (23) that the genetic hazard associated with radioiodine therapy for hyperthyroidism is so small that it should not be the major reason for avoiding radioiodine therapy.

THYROIDECTOMY Surgical therapy of hyperthyroidism has become progressively less popular in the 35 years since effective medical therapy has been available. This is partly because thyroidectomy is more expensive, both in hospital and professional costs, and results in more time lost from work than does medical therapy. Yet, in experienced hands, thyroidectomy is an effective and safe approach. Lasting alleviation of thyrotoxicosis is provided in most patients, and the preoperative use of antithyroid drugs, propranolol and iodine, greatly reduces surgical complications. The risk of permanent hypoparathyroidism and recurrent nerve deficit inversely reflects the surgeon's experience and skill while the incidence of permanent hypothyroidism is related to the amount of thyroid tissue remaining after surgery.

The use of propranolol in preparation for thyroidectomy has been described by several groups (24, 25). In a recent review of their experience, Zonszein and colleagues noted the maximal clinical response 48–72 hours after starting propranolol (26) and demonstrated the utility of this agent as the sole preoperative treatment for hyperthyroid patients.

Alteration of Peripheral Effects of Thyroid Hormones

Antithyroid drugs and radioiodine require several weeks to return hyperthyroid patients to a euthyroid state. Even with drugs that inhibit thyroid hormone secretion (e.g. iodide or lithium), improvement requires days to weeks because of the long serum half-life of thyroxine. Inhibition of the peripheral effect of thyroid hormone would provide more rapid clinical improvement. This could be accomplished by decreasing thyroid hormone binding to intracellular receptor sites or by inhibiting events subsequent to this hormone-receptor interaction. Currently, such an approach is not available.

INHIBITION OF T_4 TO T_3 CONVERSION The demonstration that much of the metabolic effect of thyroxine (T_4) results from its extrathyroidal deiodination to triiodothyronine (T_3) has provided another theoretical therapeutic approach. T_4 is converted to the more potent T_3 by microsomal 5'-deiodinases in the liver, kidney, and other organs (27). Several drugs inhibit T_4 to T_3 deiodination in both euthyroid and hyperthyroid subjects. Among the thionamides, propylthiouracil inhibits T_4 to T_3 conversion but methimazole does not (28). In hyperthyroid patients, propylthiouracil causes T_3 concentration to fall by as much as 50% in 72 hours and produces a more rapid decline in serum T_3 in combination with iodide than does methimazole (29). However, the practical significance of this phenomenon is not clear. To date, no studies have clearly demonstrated superiority of propylthiouracil over methimazole in the rapidity with which clinical features of thyrotoxicosis are ameliorated.

Propranolol (30), dexamethasone (31), and iodinated radiocontrast agents (32, 33) also decrease serum T_3 levels, apparently by peripheral inhibition of T_4 deiodination.

β-ADRENERGIC BLOCKADE It has long been recognized that many of the cardiovascular, metabolic, and neuromuscular features of hyperthyroidism resemble those of catecholamine excess. Increased β-adrenergic effects occur in hyperthyroidism in spite of normal or low plasma catecholamine concentrations (34). Some studies have demonstrated increased myocardial responsiveness to β-adrenergic agents (35). Increased catecholamine sensitivity may be related to the increased number of β-adrenergic receptor sites that have been demonstrated in hyperthyroid animals (36).

In the past, catecholamine-depleting drugs, such as reserpine and guanethidine, were used successfully to control these adrenergic symptoms. Because of its more rapid and reliable onset of action and fewer side effects, propranolol, a competitive β-adrenergic inhibitor, has replaced catecholamine-depleting drugs and has become an important adjunct in the symptomatic therapy of thyrotoxicosis. Propranolol also directly suppresses myocardial function, which may account for part of its effectiveness. Decreases in pulse rate, palpitation, sweating, lid-lag, stare, tremor, and a pronounced calming effect are seen in most patients treated with propranolol or other β-adrenergic blockers (37). Symptomatic improvement is noted within 24–48 hours after beginning therapy. Anorexia, nausea, and depression may be associated with propranolol therapy, but these side effects rarely preclude its use. Beginning doses are 40 mg orally every 6 hours, but larger doses may be required. Plasma concentrations of propranolol following oral doses are extremely variable in hyperthyroid patients (38), and the dose must be adjusted to control the symptoms of the individual patient. The pulse rate and cardiovascular symptoms are the most reliable parameters to follow. Propranolol is generally contraindicated in patients with a history of asthma or with congestive heart failure. However, β-blockers are useful in hyperthyroid patients whose heart failure is aggravated by tachycardia (39). Cautious administration of propranolol, in a carefully monitored setting, is indicated in such patients.

Ancillary Therapeutic Measures

Rest and sedation were former mainstays in the management of anxious, hyperkinetic, hyperthyroid patients. As has been mentioned, propranolol is usually effective in rapidly improving neuropsychiatric aspects of thyrotoxicosis. However, mild sedation is occasionally necessary, especially during the night.

Nutritional aspects of hyperthyroidism are not often emphasized. Patients with long-standing or severe hyperthyroidism are usually in negative protein and calorie balance. A high caloric diet with multivitamin supplementation is appropriate for these patients until their thyroid function begins to normalize. Hypercalciuria and negative calcium balance are common features of hyperthyroidism (40). Hypercalcemia is also occasionally seen (41). Unless serum calcium is elevated, a high calcium intake is indicated.

Therapy of Nonthyroid Problems in Hyperthyroid Patients

Patients with hyperthyroidism may present with or develop other medical or surgical problems in the course of their hyperthyroidism. Management of medical problems may be complicated by alterations in drug absorption or metabolism in thyrotoxicosis. Metabolism of many drugs is enhanced, serum half-lives are shortened, and the dose required to achieve a therapeutic serum level may be increased. This has been clearly demonstrated in the case of digoxin therapy (42). Careful monitoring of serum concentrations or of the patient's response to pharmaceutical compounds is extremely important. An acute surgical problem in a hyperthyroid patient can best be managed by controlling the patient's cardiovascular status preoperatively with propranolol (26).

THERAPY OF SPECIFIC TYPES OF HYPERTHYROIDISM

Graves' Disease

Graves' disease is classically defined as a triad of a diffuse goiter, hyperthyroidism, and infiltrative ophthalmopathy. A unique infiltrative disorder of the skin, pretibial myxedema, is a less commonly associated feature. Hyperthyroidism occurs independently of the other features and is the principal focus of this discussion. Although the basic pathophysiologic defect in Graves' disease is unknown, an autoimmune mechanism is most probable. Circulating thyroid-stimulating immunoglobulins and thyroid-sensitized T-lymphocytes, as demonstrated by newer techniques, are present in virtually all of these patients. Several excellent reviews have recently discussed these etiologic factors in detail (43).

ADULTS Graves' disease is characterized by periods of thyroid hyperactivity followed by spontaneous remission in many patients. This feature permits the successful therapy of these patients with a course of antithyroid drug therapy that controls the hyperthyroidism until spontaneous remission occurs. There is no clear evidence that thionamides alter the course of the autoimmune process, although immunosuppressive effects of propylthiouracil (PTU) have been demonstrated in vitro (44). Of the principal modalities employed in the treatment of thyrotoxicosis, antithyroid drugs are the least likely to induce permanent remission. Nevertheless, they are the most commonly used form of therapy for Graves' disease in many medical centers, including our own, for two important reasons: they do not produce permanent morbidity (hypothyroidism, hypoparathyroidism, etc.), and they do not require special skills, training, or licensing, and thus can be prescribed by general physicians. Patients of any age from birth to senility can be successfully treated with antithyroid drugs. However, since the thyroid-blocking effect of thionamides wears off within 2–4 days after interrupting therapy, compliance with the prescribed regimen is necessary.

Euthyroidism is usually achieved within 3–4 months of instituting thionamide therapy. If improvement in clinical or laboratory findings is not noted within 6 weeks of instituting therapy, the dose is doubled. In patients who have not improved after 3 months, an alternative form of therapy is chosen rather than allowing them to remain hyperthyroid. Fortunately, this happens rarely.

Recently it has been suggested that the incidence of sustained remissions of hyperthyroidism after antithyroid drug therapy in the United States has gradually decreased, perhaps related to an increase in iodine intake (45). However, our recent experience (46) and that of the Lahey Clinic (47) suggest that remission rates of 30–40% are still seen after thionamide therapy, and there is evidence that an increase in iodine intake does not cause a decreased lasting remission rate (48, 49). Unfortunately, there is no way of predicting with certainty which patients will remain in remission after antithyroid drugs are discontinued. In most series, patients with a small goiter and relatively mild hyperthyroidism are the most successfully managed. Various tests, including suppression of radioiodine or pertechnetate thyroid uptake by T_3 (50, 51) and measurement of thyroid-stimulating immunoglobulin (52), have been proposed to identify patients who have undergone spontaneous remission while on antithyroid drugs. However, these tests are cumbersome to perform or are not routinely available, and none of them is completely reliable. Until the technology is significantly improved, we think it is preferable to use the cheaper and more reliable test of simply stopping therapy after the appropriate time and observing the patient's course.

It has been standard practice to treat patients for 1–2 years before discontinuing thionamide therapy. Slingerland & Burrows (53) reported their experience with long-term antithyroid drug therapy. They continued therapy until 24-hour ^{131}I uptake was suppressed to less than 20% by desiccated thyroid or T_3. This required an average of 4.4 years (range 1–14), and the remission rate was 67%. However, 20% of patients with abnormal suppression tests stayed in remission for more than a year, suggesting that, with this approach, treatment of some patients is prolonged unnecessarily.

A lasting remission rate of 39% in patients who were treated only until they became euthyroid within an average of 4.5 months has been reported (46). Until a reliable method of predicting which patients will remain in remission is available, we think that an initial short course of antithyroid drugs is justified. The major advantage of short-term therapy is that the cost and inconvenience to the patient is minimized in those who remain in lasting remission after discontinuing the drug. For patients who relapse after short-term antithyroid therapy, a longer course of antithyroid drugs or ablative therapy is appropriate.

Radioiodine provides a simple approach to therapy and is probably the most commonly employed agent in patients with Graves' disease. Improvement is generally noted in 4–8 weeks and euthyroidism is achieved within 3–4 months following treatment. Therapy is almost always performed as an outpatient procedure. Exacerbation of hyperthyroidism following ^{131}I therapy occasionally occurs (54). However, in most patients thyroid hormone levels change only minimally after ^{131}I administration (55). Pretreatment of patients with antithyroid drugs has been suggested to prevent radiation-induced thyroid storm, but it has not been proven effective. Initiation of propranolol prior to ^{131}I administration is indicated for symptomatic relief in many patients and minimizes cardiovascular effects of radiation thyroiditis.

The major complication of ^{131}I therapy is hypothyroidism. A biphasic pattern in the development of hypothyroidism following ^{131}I therapy has been observed by several investigators (56). The incidence one year after therapy is roughly related to the dose of radioiodine administered. Thereafter, a progressive increase of 2–5% per year is seen regardless of the initial dose. Not surprisingly, reported incidences of hypothyroidism vary considerably depending upon the initial dose and the duration of follow-up. The appropriate dose of ^{131}I would be that with which the smallest number of patients remain hyperthyroid and the least number develop hypothyroidism. The most commonly used doses are 80–100 μCi per gram of thyroid tissue. Various regimens have been devised for calculating the dose based upon an estimate of thyroid gland size and the concentration of isotope in the gland as determined by radioiodine uptake. Adequate comparison of these regimens has not been made. Roudebush and colleagues (57) described a reasonable approach in which the delivered dose of ^{131}I per gram of thyroid tissue is increased in patients with large goiters since such thyroids are more resistant to radioiodine.

Propranolol alone has been advocated as the primary therapy in patients with Graves' disease. Lowe and co-workers (58) reported remissions of more than 18 months in 32% of patients so treated. It was impossible to predict which patients would undergo remission during the first year of therapy. Because β-blockade does not correct the hypermetabolism or enhanced myocardial contractility in hyperthyroidism, we feel that adrenergic blockade should be used as an adjunct to antithyroid therapy rather than as the sole approach to management.

CHILDREN Graves' disease in children and adolescents closely resembles the same disease in adults and the approach to therapy is similar. Antithyroid drugs are used initially. Compliance with drug therapy is especially important in younger patients, and the administration of the drug in a single daily dose is of particular advantage in these patients. The remission rates with antithyroid drugs in children are similar to those seen in adults (59). Patients who relapse after discontinuation of antithyroid drugs are managed with chronic antithyroid therapy, surgery or radioiodine. The use of radioiodine in children and adolescents remains controversial because of concern about malignancy and genetic damage. Nevertheless, ^{131}I has been used with impressive results in some centers as the primary therapy in these patients (60). Hyperthyroidism was controlled in an average of 3.3 months. No complications other than hypothyroidism were observed during a follow-up period averaging 12 years. Although a full appraisal of the complications of radioiodine therapy in children awaits longer term follow-up, we feel that these risks are less than those associated with surgical therapy, and we do not hesitate to use ^{131}I in children if antithyroid drugs are not effective.

Regardless of the therapy employed, the management of Graves' disease requires the establishment of a firm patient-physician relationship and education of the patient about the disease. Patient understanding of and compliance with the therapeutic program and a systematic schedule of evaluation is necessary during the treatment of the hyperthyroidism. Life-long follow-up of thyroid function is required for all patients with Graves' disease. Relapse of hyperthyroidism many years after discontinuation of drug therapy has been described, and the late development of hypothyroidism following ablative therapy or in patients treated only with antithyroid drugs or potassium iodide (61, 62) has been documented. Patience and persistence on the part of both the patient and the physician are key ingredients in a successful therapeutic program.

CONGENITAL GRAVES' DISEASE This is a rare condition in which infants of mothers with active or previously active Graves' disease are born with neonatal thyrotoxicosis. Although there is evidence that congenital Graves' disease is caused by transplacental migration of thyroid-stimulating immunoglobulin (TSI) from the mother to the fetus (63), a detailed analysis by Hollingsworth & Mabry (64) indicates that in many of these children the disease persists for months or years, much longer than could be explained by a persistence of TSI passively transferred from the mother before birth. It is therefore probable that some of these infants have active rather than passive Graves' disease.

Conventional management includes the use of antithyroid drugs and glucocorticoids (65). Propranolol has been used successfully to stabilize the cardiovascular status in some patients (66). Antithyroid drugs should be continued until the patient becomes euthyroid. Because the condition is so rare, few data are available on the most appropriate therapy.

OPHTHALMOPATHY There is still no satisfactory treatment for the severe ophthalmopathy of Graves' disease. The evaluation of therapeutic programs is made difficult by the occurrence of spontaneous exacerbations and remissions of the eye disease. Fortunately, Graves' ophthalmopathy is usually mild, is infrequently progressive (67), and may resolve completely.

Although the infiltrative ophthalmopathy is not due to hyperthyroidism, normalization of thyroid function usually improves the lid retraction and ophthalmoplegia. In severe or progressive cases, high-dose corticosteroid therapy may decrease the inflammatory and infiltrative manifestations. Recent experience with cyclophosphamide in such patients is encouraging (68) and deserves further evaluation. Kriss and co-workers (69) reported favorable results in patients with severe ophthalmopathy who received retrobulbar irradiation with a linear accelerator early in the course of their disease. However, long-term results or the experience of others using this technique are not yet available.

General supportive measures such as methylcellulose eyedrops, eye patches, or tarsorrhaphy may relieve pain and burning and minimize the danger of corneal ulceration. If vision is threatened, surgical orbital decompression may be necessary. Extraocular muscle surgery to relieve diplopia or purely cosmetic surgery should be postponed until the active phase of the disease has disappeared. Progressive changes in the eye muscles often occur during the active phase of the disease, making reoperation necessary if surgery is performed prematurely.

DERMOPATHY Localized "myxedema" is a rare complication of Graves' disease which usually occurs in the pretibial area and resembles the myxedema of hypothyroidism only in that there is a subcutaneous infiltration of mucopolysaccharides and nonpitting swelling in the involved skin area. The cause is unknown, although it occurs only in patients with ophthalmopathy (with or without hyperthyroidism). It is not due to an inability of thyroid hormone to reach the involved area (70), but it may be due to an abnormal serum fibroblast-stimulating factor (71). Local glucocorticoid therapy often provides symptomatic relief from the burning and itching that may accompany the dermopathy.

Autonomous Functioning Nodular Goiter

The distinctions between autonomously functioning nodular goiter (toxic nodular goiter) and Graves' disease, which were first made by Plummer more than 65 years ago, are still pertinent. Unlike Graves' disease, in which diffuse thyroid hyperplasia and hyperfunction is observed, toxic nodular goiter is characterized by functionally heterogeneous populations of thyroid follicles (72). Normal follicles retain their TSH-dependence while other follicles secrete thyroid hormone independent of the normal homeostatic control of thyroid function. Hyperfunctioning follicles may be localized in a discrete solitary nodule or may be dispersed throughout a multinodular or diffusely enlarged thyroid gland. The degree of thyroid hormone overproduction depends upon the mass of autonomously functioning follicles and the efficiency with which they function. The clinical manifestations of hyperthyroidism in patients with autonomously functioning nodular goiters are indistinguishable from those seen in Graves' disease except for the absence of infiltrative ophthalmopathy and dermopathy in the former. The generalities of therapeutic approaches to the two diseases are similar. Many of the cardiovascular and neuromuscular features of toxic nodular goiter can be controlled by propranolol. Antithyroid drugs effectively block the synthesis of thyroid hormone and euthyroidism can be achieved. However, spontaneous remissions are infrequently seen in toxic nodular goiter, and hyperthyroidism almost always recurs after discontinuing antithyroid therapy. Consequently the thionamides are generally used only in the preparation of patients for surgical or radioiodine therapy and only rarely for a definitive attack.

Solitary autonomously functioning thyroid nodules may be associated with normal thyroid function or with hyperthyroidism. In general, thyroid hormone production is related to the size of the nodule. Nodules that produce hyperthyroidism are usually over 2.5 cm in diameter. Although evolution from a small autonomous nodule in a euthyroid patient to a larger nodule causing hyperthyroidism has been described, this transition occurs much less frequently than one would expect (73). Treatment of a solitary autonomous nodule is generally not necessary unless hyperthyroidism exists. Toxic solitary nodules can be successfully treated with either radioiodine or surgery. The latter provides prompt control and operative complications are much less common than in patients with Graves' disease since the dissection is restricted to one lobe. Transient suppression of the pituitary-thyroid axis and a delay in the return of function in the contralateral lobe has been observed post-operatively, but permanent hypothyroidism is rarely reported.

Various dosage regimens of ^{131}I have been used to treat solitary toxic nodules. In general, larger doses than those used for Graves' disease are employed. Administration of 15–25 mCi to the nodule cured the hyperthyroidism in 16 of 17 patients reported by Hamburger (73). In that series, the nodules decreased in size after radioiodine therapy but remained palpable. In the largest series reported (74), doses calculated to deliver 30,000 rads to the nodule controlled the hyperthyroidism in 80% of the patients. In nodules greater than 5 cm in diameter, this may require an administered dose of ^{131}I in excess of 100 mCi (75), resulting in significant gonadal and whole-body radiation. Post-radioiodine hypothyroidism is seen infrequently in these patients since the contralateral lobe is inactive.

Toxic multinodular goiter generally occurs in patients over the age of 50 who have a long history of preexisting goiter. Radioiodine is most often used in these patients. In patients with cardiovascular disease, pretreatment with antithyroid drugs and/or propranolol is advised, although there is little evidence to document the efficacy of this therapy in minimizing exacerbation of hyperthyroidism after radioiodine Initial doses of 30–50 mCi are given and multiple doses are often necessary. Subtotal thyroidectomy, after control of hyperthyroidism with antithyroid drugs, is recommended in patients whose goiter produces compressive symptoms.

Subacute Thyroiditis

Subacute thyroiditis, a presumed viral disease, produces inflammation of the thyroid with pain, tenderness, and swelling of the gland accompanied by malaise and low-grade fever (76). Release of hormone from the disrupted follicles results in transient hyperthyroidism that lasts less than 6 weeks. Thyroid radioiodine uptake is low during the hyperthyroid phase of the illness. "Silent," nonpainful, subacute thyroiditis has been described as a cause of hyperthyroidism (77), but almost all such cases are probably really lymphocytic thyroiditis. Salicylates usually control the inflammatory symptoms. In patients with severe symptoms, corticosteroids are often used. Since the thyroid is not hyperfunctional, antithyroid drugs, radioiodine, or surgical therapy are not indicated. Propranolol may be employed if cardiovascular manifestations are marked.

Lymphocytic Thyroiditis

Several recent reports have described patients with transient, sometimes recurrent, hyperthyroidism with minimal or no inflammatory manifestations and low radioiodine uptake (78–80). Thyroid biopsies in these patients have demonstrated lymphocytic infiltration and follicular disruption. The etiology of this disorder and its relationship to Hashimoto's thyroiditis are not known. Of special note is the importance of differentiating this disorder from Graves' disease. This is best done by measuring radioiodine uptake while the patient is hyperthyroid. Propranolol is the only therapy required in these patients.

Patients may have coexisting Graves' disease and lymphocytic thyroiditis ("Hashitoxicosis") (81). Treatment is identical to that for Graves' disease.

Trophoblastic Disease

Patients with hydatidiform moles or with choriocarcinoma may present with hyperthyroidism and a diffuse goiter resembling Graves' disease (82). Human chorionic gonadotrophin (hCG) is structurally related to TSH, has weak thyroid-stimulating activity, is found in very high concentration in these patients, and is thought to be responsible for the thyroid hyperactivity (83). Chemotherapy for choriocarcinoma or evacuation of the mole results in prompt resolution of the hyperthyroidism as hCG levels fall. Propranolol can control cardiovascular manifestations until the thyrotoxicosis abates.

Hyperfunctioning Thyroid Carcinoma

Follicular carcinoma is a very uncommon cause of hyperthyroidism (84). Because the efficiency with which follicular carcinomas synthesize thyroid hormone is less than that of normal tissue, hyperthyroidism is generally seen only in patients with a very large tumor or with widespread metastases. This clinical situation should not be confused with the patient who presents with a hyperfunctioning thyroid nodule confined to the neck without evidence of metastasis. Although thyrotoxicosis can be controlled with antithyroid drugs and propranolol, the approach to these patients is surgical excision of accessible tumor followed by administration of very large doses of ^{131}I to ablate the remaining malignant tissue (85).

SPECIAL THERAPEUTIC PROBLEMS

Thyroid Storm

Thyroid storm is a life-threatening problem in which there is hyperprexia, delirium, extreme tachycardia, and agitation. Although full-blown thyroid storm is rare, severe thyrotoxicosis with a potential danger of storm is seen frequently. Precipitating factors are varied and include acute infections, surgery (thyroidal or extrathyroidal), other medical illnesses, iodine administration (86), and radioiodine therapy (54).

Thyroid storm has frequently been assumed to be caused by a sudden marked increase in circulating thyroid hormone, such as might be expected from trauma to the thyroid during thyroidectomy. However, in those few studies that have been performed, there is no evidence that storm is associated with an increase in circulating thyroid hormone. The common background is an acute major stress. Patients with hyperthyroidism are more sensitive to catecholamines than are euthyroid individuals. Since acute stress results in a major increase in adrenergic effects, it is probable that increased sensitivity to a surge in catecholamines is the major etiologic factor.

Thyroid storm is a major catastrophe that resolves in a few days, usually in recovery but sometimes fatally. Some of the standard therapeutic measures therefore appear ill conceived. Huge doses of thionamide drugs, intravenous iodide, etc, are of little practical or theoretical value because it takes days or weeks for them to have any significant effect. Since current evidence indicates that the primary cause of thyroid storm is a marked increase in β-adrenergic effect, an appropriate β-adrenergic antagonist, such as propranolol, seems most appropriate. Propranolol is extremely effective in treating thyroid storm, but it usually has to be given in large doses, 20–80 mg or more every 4 hours orally. In patients unable to take oral medication, 1–5 mg may have to be given intravenously. Lysis of the storm is usually achieved within a few hours with the use of propranolol (87).

Glucocorticoids have been routinely used in thyroid storm to correct a presumed relative deficiency of adenocortical hormones. While there are no data supporting such a deficiency, the acute inhibition of thyroid hormone secretion by dexamethasone in Graves' disease (88) may be a useful therapeutic adjunct.

In patients who do not respond to conventional therapy of storm, plasmapheresis and peritoneal dialysis have been successfully employed to remove plasma proteins with their associated high concentration of bound thyroid hormone (89, 90).

General supportive measures are required, including physical cooling of the patient and correction of hypoxemia and fluid and electrolyte abnormalities. Salicylates are contraindicated because they displace thyroid hormone from serum-binding protein and thereby increase free thyroid hormone concentrations (91). Precipitating and associated medical problems should be managed simultaneously.

As soon as the patient is past the crisis stage (usually 1–3 days), appropriate therapy for the underlying thyrotoxicosis can be instituted. Such therapy can be chosen on the basis of the etiology of the thyrotoxicosis and the various factors considered important in managing it, but should not include emergency decisions that may lead to unfortunate delays in definitive therapy. For instance, large doses of iodide should not be given to patients for whom radioiodine therapy is planned.

Pregnancy

Antithyroid drugs are the treatment of choice in thyrotoxic pregnant women. Radioiodine is obviously contraindicated because of the danger of fetal thyroid damage. Surgery carries a greater risk than antithyroid drugs.

Since thionamides are transferred across the placenta, fetal goiter and hypothyroidism can occur, and the dose is kept to a minimum during pregnancy: 15 mg methimazole or 150 mg propylthiouracil daily during the last trimester is usually sufficient to keep the maternal hyperthyroidism under adequate control, and fetal complications are extremely rare with these doses (92). Fetal hypothyroidism cannot be prevented by maternal administration of thyroid hormones because these traverse the placenta poorly.

For the unusual pregnant patient with severe hyperthyroidism who cannot be controlled with low doses of antithyroid drugs, subtotal thyroidectomy during the second trimester is indicated. Although propranolol has been used without complication in the treatment of hyperthyroidism during pregnancy (93), β-adrenergic blockade compromises placental blood flow and should be used cautiously. If antithyroid drugs are continued postpartum, the mother should not breast-feed the infant since antithyroid drugs may be secreted in the milk.

Literature Cited

1. Tunbridge, W. M. G., Evered, D. C., Hall, R., Appleton, D., Brewis, M., Clark, F., Evans, J. G., Young, E., Bird, T., Smith, P. A. 1977. The spectrum of thyroid disease in a community: the Wickham survey. *Clin. Endocrinol.* 7:481–93
2. Bürgi, H., Geiser, J., Rösler, H., Studer, H. 1978. Die verkannte Hyperthyreose beim Spitalpatienten. *Schweiz. Med. Wochenschr.* 108:1257–62
3. Langer, P., Greer, M. A. 1977. *Antithyroid Substances and Naturally Occurring Goitrogens.* Basel: Karger. 178 pp.
4. Richards, J. B., Ingbar, S. H. 1959. The effects of propylthiouracil and perchlorate on the biogenesis of thyroid hormone. *Endocrinology* 65:198–207
5. Iino, S., Yamada, T., Greer, M. A. 1961. Effect of graded doses of propylthiouracil on biosynthesis of thyroid hormones. *Endocrinology* 68:582–88
6. Stanley, M. M., Astwood, E. B. 1959. 1-methyl-2-mercaptoimidazole: An antithyroid compound highly active in man. *Endocrinology* 44:588–89
7. Greer, M. A., Meihoff, W. C., Studer, H. 1965. Treatment of hyperthyroidism with a single daily dose of propylthiouracil. *N. Engl. J. Med.* 272:888–91
8. Kammer, H., Srinavasan, K. 1969. The use of antithyroid drugs in a single daily dose; treatment of diffuse toxic goiter. *J. Am. Med. Assoc.* 209:1325–27
9. Gwinup, G. 1978. Prospective randomized comparison of propylthiouracil. *J. Am. Med. Assoc.* 239:2457–59
10. Rosove, M. H. 1977. Agranulocytosis and antithyroid drugs. *West. J. Med.* 126:339–43
11. Grainger, A., Gregson, D. A. 1945. Thiouracil in the treatment of thyrotoxicosis. *Br. Med. J.* 2:343–45
12. Trotter, W. R. 1962. The relative toxicity of antithyroid drugs. *J. New Drugs* 2:333–43
13. Kasai, K., Suzuki, H., Kikuchi, T., Ieiri, T., Takemura, Y., Shimoda, S. I. 1976. Clinical study on early changes in thryoid function in hyperthyroidism treated with propylthiouracil and a relatively small dose of iodide. *Endocrinol. Jpn.* 23:347–54
14. Schimmel, M., Utiger, R. D. 1977. Acute effect of inorganic iodide after ^{131}I therapy for hyperthyroidism. *Clin. Endocrinol.* 6:329–32
15. Schou, M., Amdisen, A., Jensen, S. E., Olson, T. 1968. Occurrence of goitre during lithium treatment. *Br. Med. J.* 3:710–13
16. Berens, S. C., Bernstein, R. S., Robbins, J., Wolff, J. 1970. Antithyroid effects of lithium. *J. Clin. Invest.* 49:1357–67
17. Temple, R., Berman, M., Robbins, J., Wolff, J. 1972. The use of lithium in the treatment of hyperthyroidism. *J. Clin. Invest.* 51:2746–56
18. Linquette, M., Lefebvre, J., Van Parys, C., Wemeau, J. L. 1978. Le lithium dans le traitement des thyrotoxicoses. *Ann. Endocrinol. (Paris)* 39:15–21
19. Kristensen, O., Harrestrup-Anderson, H., Pallisgaard, G. 1976. Lithium carbonate in the treatment of thyrotoxicosis. *Lancet* 1:603–5
20. Turner, J. C., Brownlie, B. E. W., Sadler, W. A., Jenson, C. H. 1976. An evaluation of lithium as an adjunct to carbimazole treatment in acute thyrotoxicosis. *Acta Endocrinol.* 83:86–92
21. McDougall, I. R., Greig, W. R. 1976. ^{125}I therapy in Graves' disease. *Ann. Intern. Med.* 85:720–23
22. Dobyns, B. M., Sheline, G. E., Workman, J. B., Tompkins, E. A., McConahey, W. M., Becker, D. V. 1974. Malignant and benign neoplasms of the thyroid in patients treated for hyperthyroidism: a report of the Cooperative Thyrotoxicosis Therapy Follow-up Study. *J. Clin. Endocrinol. Metab.* 38:976–98

23. Robertson, J. S., Gorman, C. A. 1976. Gonadal radiation dose and its genetic significance in radioiodine therapy of hyperthyroidism. *J. Nucl. Med.* 17: 826–35
24. Bewsher, P. D., Pegg, C. A. S., Stewart, D. J. 1974. Propranolol in the surgical management of thyrotoxicosis. *Ann. Surg.* 180:787–90
25. Toft, A. D., Irvine, W. J., McIntosh, D., MacLeod, D. A. D., Seth, J., Cameron, E. H. D., Lidgard, G. P. 1976. Propranolol in the treatment of thyrotoxicosis by subtotal thyroidectomy. *J. Clin. Endocrinol. Metab.* 43:1312–16
26. Zonszein, J., Santangelo, R. P., Mackin, J. F., Lee, T. C., Coffey, R. J., Canary, J. J. 1979. Propranolol therapy in thyrotoxicosis. *Am. J. Med.* 66:411–16
27. Braverman, L. E., Ingbar, S. H., Sterling, K. 1970. Conversion of thyroxine (T_4) to triiodothyronine (T_3) in athyreotic human subjects. *J. Clin. Invest.* 49:855–64
28. Oppenheimer, J. H., Schwartz, H. L., Surks, M. I. 1972. Propylthiouracil inhibits the conversion of L-thyroxine to L-triiodothyronine. *J. Clin. Invest.* 51:2493–97
29. Abuid, J., Larsen, P. R. 1974. Triiodothyronine and thyroxine in hyperthyroidism. Comparison of the acute changes during therapy with antithyroid agents. *J. Clin. Invest.* 54:201–9
30. Tevaarwerk, G. J. M., Malik, M. H., Boyd, D. 1978. Preliminary report on the use of propranolol in thyrotoxicosis: 1. Effect on thyroxine, triiodothyronine and reverse triiodothyronine concentrations. *Can. Med. Assoc. J.* 119:350–60
31. Chopra, I. J., Williams, D. E., Orgiazzi, J., Solomon, D. H. 1975. Opposite effects of dexamethasone on serum concentrations of 3,3',5' triiodothyronine (reverse T_3), and 3,3',5 triiodothyronine (T_3). *J. Clin. Endocrinol. Metab.* 41:911–20
32. Bürgi, H., Wimpfheimer, C., Burger, A., Zaunbauer, W., Rösler, H., Lemarchaud-Beraud, T. 1976. Changes of circulating thyroxine, triiodothyronine, and reverse triiodothyronine after radiographic contrast agents. *J. Clin. Endocrinol. Metab.* 43:1203–10
33. Burger, A., Dinichert, D., Niced, P., Jenny, M., Lemarchand-Beraud, T., Valloton, M. B. 1976. Effect of aminodarone on serum triiodothyronine, reverse triiodothyronine, thyroxine, and thyrotropin: a drug influencing metabolism of thyroid hormones. *J. Clin. Invest.* 58:255–59
34. Coulombe, P., Dussault, J. H., Walker, P. 1976. Plasma catecholamine concentrations in hyperthyroidism and hypothyroidism. *Metabolism* 25:973–80
35. Wildenthal, K. 1972. Studies of isolated fetal mouse hearts in organ culture. *J. Clin. Invest.* 51:2702–9
36. Williams, L. T., Lefkowitz, R. J., Watanabe, A. M., Hathaway, D. R., Besch, H. R. Jr. 1977. Thyroid hormone regulation of beta-adrenergic receptor number: possible biochemical basis for the hyperadrenergic state in hyperthyroidism. *Clin. Res.* 25:458A (Abstr.)
37. Grossman, W., Robin, N. I., Johnson, L. W., Brooks, H., Selenkow, H. A., Dexter, L. 1971. Effects of beta blockade on the peripheral manifestations of thyrotoxicosis. *Ann. Intern. Med.* 74:875–79
38. Rubenfeld, S., Silverman, V. E., Welch, K. M. A., Mallette, L. E., Kohler, P. O. 1979. Variable plasma propranolol levels in thyrotoxicosis. *N. Engl. J. Med.* 300:353–54
39. Weiner, L., Stout, B. D., Cox, J. W. 1969. Influence of beta-sympathetic blockade (propranolol) on the hemodynamics of hyperthyroidism. *Am. J. Med.* 46:227–33
40. Smith, D. A., Frazer, S. A., Wilson, G. M. 1973. Hyperthyroidism and calcium metabolism. *Clin. Endocrinol. Metab.* 2:333–36
41. Maxon, H. R., Apple, D. J., Goldsmith, R. E. 1978. Hypercalcemia in thyrotoxicosis. *Surg. Gynecol. Obstet.* 147: 694–96
42. Croxson, M. S., Ibbertson, H. K. 1975. Serum digoxin in patients with thyroid disease. *Br. Med. J.* 3:566–68
43. Solomon, D. H., Kleeman, K. E. 1976. Concepts of pathogenesis of Graves' disease. *Adv. Intern. Med.* 22:273–99
44. Wall, J. R., Manwer, G. L., Greenwood, D. M., Walters, B. A. 1976. The in vitro suppression of lectin induced ^{3}H-thymidine incorporation into DNA of peripheral blood lymphocytes after the addition of propylthiouracil. *J. Clin. Endocrinol. Metab.* 43:1406–9
45. Wartofsky, L. 1973. Low remission after therapy for Graves' disease. *J. Am. Med. Assoc.* 226:1083–88
46. Greer, M. A., Kammer, H., Bouma, D. A. 1977. Short-term antithyroid drug therapy for the thyrotoxicosis of Graves's disease. *N. Engl. J. Med.* 297:173–76
47. Totten, M. A., Wool, M. S. 1979. Medical treatment of hyperthyroidism. *Med. Clin. North Am.* 63:321–28

48. Lumholtz, I. B., Poulsen, D. L., Siersboek-Neilsen, K., Friis, T., Rogowski, P., Kirkegaard, C., Molholm Hansen J. 1977. Outcome of long-term antithyroid treatment of Graves' disease in relation to iodine intake. *Acta Endocrinol.* 84:538–41
49. Nagataki, S., Uchimura, H., Ikeda, H., Maeda, M., Kuzuya, N., Chiu, S. C., Kumagai, L. F. 1976. Effect of chronic iodide ingestion on the incidence of relapse in T_3-suppressible patients treated with thionamide drugs. *Acta Endocrinol. Suppl.* 204:76 (Abstr.)
50. Van'T Hoff, W., Pover, G. G., Drury, P. L. 1978. Technetium uptake in the management of thyrotoxicosis. *Clin. Endocrinol.* 8:327–33
51. Yamamoto, M., Igarashi, T., Kimura, S., Tsukamoto, S., Togawa, K., Ogata, E. 1979. Thyroid suppression test and outcome of hyperthyroidism treated with antithyroid drugs and triiodothyronine. *J. Clin. Endocrinol. Metab.* 48:72–77
52. Davies, T. F., Evered, D. C., Rees Smith, B., Yeo, P. P. B., Clark, F., Hall, R. 1977. Value of thyroid-stimulating-antibody determinations in predicting short-term thyrotoxic relapse in Graves' disease. *Lancet* 1:1181–82
53. Slingerland, D. W., Burrows, B. A. Results of long-term drug treatment of hyperthyroidism. 1978. *Program 54th Meet. Am. Thyroid Assoc.* T1 (Abstr.)
54. Shafer, R. B., Nuffal, F. Q. 1971. Thyroid crisis induced by radioiodine. *J. Nucl. Med.* 12:262–64
55. Tamagna, E. I., Levine, G. A., Hershman, J. M. 1979. Thyroid hormone concentrations after radioiodine therapy for hyperthyroidism. *J. Nucl. Med.* 20:387–91
56. Becker, D. V., McConahey, W. M., Dobyns, B. M., Tompkins, E., Sheline, G. E., Workman, J. B. 1971. The results of radioactive iodine treatment of hyperthyroidism: a preliminary report of the Thyrotoxicosis Therapy Follow-up Study. In *Further Advances in Thyroid Research,* Vol. 1, ed. K. Fellinger, R. Hofer, pp. 603–9. Vienna: Wiener Medizinischen Akademie. 632 pp.
57. Roudebush, C. P., Hoye, K. E., DeGroot, L. J. 1977. Compensated low-dose ^{131}I therapy of Graves' disease. *Ann. Intern. Med.* 87:441–43
58. Lowe, D. C., Hadden, D. R., Montgomery, D. A. D., Weaver, J. A., 1975. Propranolol as the sole therapy for thyrotoxicosis: Long-term follow up. In *Thyroid Research,* ed. J. Robbins, L. E. Braverman, pp. 429–33. Amsterdam: Excerpta Medica. 638 pp.
59. Vaida, V. A., Bongiovanni, A. M., Parks, J. S., Tenore, A., Kirkland, R. T. 1974. Twenty-two years' experience with medical management of juvenile thyrotoxicosis. *Pediatrics* 54:565–70
60. Safa, A. M., Schumacher, P. O., Rodriguez-Antunez, A. 1975. Long-term follow-up results in children and adolescents treated with radioactive iodine (^{131}I) for hyperthyroidism. *N. Engl. J. Med.* 292:167–71
61. Wood, L. C., Peterson, M., Ingbar, S. H. 1972. Delayed thyroid failure and decreased reserve following antithyroid therapy in Graves' disease. *Program 48th Meet. Am. Thyroid Assoc.* 61 (Abstr.)
62. Wood, L. C., Maloof, F. 1975. Thyroid failure after potassium iodide treatment of diffuse toxic goiter. *Trans. Assoc. Am. Physicians* 88:235–47
63. Dirmikis, S. M., Munro, D. S. 1975. Placental transmission of thyroid-stimulating immunoglobulins. *Br. Med. J.* 2:665–66
64. Hollingsworth, D. R., Mabry, C. 1976. Congenital Graves' disease. *Am. J. Dis. Child.* 130:148–55
65. Fisher, D. A. 1976. Pathogenesis and therapy of neonatal Graves' disease. *Am. J. Dis. Child.* 130:133–34
66. Smith, C. S., Howard, N. J. 1973. Propranolol in treatment of neonatal thyrotoxicosis. *J. Pediatr.* 83:1046–48
67. Solem, J. H., Segaard, E., Ytteborg, J. 1978. The course of endocrine ophthalmopathy during antithyroid therapy in a prospective study. *Acta Med. Scand.* 205:111–14
68. Bigos, S. T., Nisula, B. C., Daniels, G. H., Eastman, R. C., Johnston, H. H., Kohler, P. O. 1979. Cyclophosphamide in the management of advanced Graves' ophthalmopathy. *Ann. Intern. Med.* 90:921–23
69. Donaldson, S. S., Bagshaw, M. A., Kriss, J. P., 1973. Supervoltage orbital radiotherapy for Graves' ophthalmopathy. *J. Clin. Endocrinol. Metab.* 37:276–85
70. Greer, M. A. 1957. Exophthalmos and localized pretibial myxedema in a euthyroid patient: Studies with triiodothyronine. *J. Clin. Endocrinol. Metab.* 17:1466–71
71. Cheung, H. S., Nimni, M. E., Kamiel, M. B., Nicoloff, J. T. 1977. Evidence for a fibroblast stimulating factor(s) in serum of patients with pretibial myxedema. *Clin. Res.* 25:147A (Abstr.)

72. Studer, H., Hunziker, H. R., Ruchti, C. 1978. Morphologic and functional substrate of thyrotoxicosis caused by nodular goiters. *Am. J. Med.* 65:227–34
73. Hamburger, J. I. 1975. Solitary autonomously functioning thyroid lesions. *Am. J. Med.* 58:740–48
74. Horst, W., Rosler, H., Schneider, C., Labhart, A. 1967. 306 cases of toxic adenoma. *J. Nucl. Med.* 8:515–28
75. Gorman, C. A., Robertson, J. S. 1978. Radiation dose in the selection of ^{131}I or surgical treatment for toxic thyroid adenoma. *Ann. Intern. Med.* 66:760–88
76. Hamburger, J. I. 1974. Subacute thyroiditis: diagnostic difficulties and simple treatment. *J. Nucl. Med.* 15:81–89
77. Papapetrou, P. D., Jackson, I. M. D. Thyrotoxicosis due to "silent" thyroiditis. *Lancet* 1:361–63
78. Gluck, F. B., Nusynowitz, M. L., Plymate, S. 1975. Chronic lymphocytic thyroiditis, thyrotoxicosis, and low radioiodine uptake. *N. Engl. J. Med.* 293:624–28
79. Woolf, P. D., Daly, R. 1976. Thyrotoxicosis with painless thyroiditis. *Am. J. Med.* 60:73–79
80. Gorman, C. A., Duick, D. S., Woolner, L. B., Wahner, H. W. 1978. Transient hyperthyroidism in patients with lymphocytic thyroiditis. *Mayo Clin. Proc.* 53:359–65
81. Fatourechi, V., McConahey, W. M., Woolner, L. B. 1971. Hyperthyroidism associated with histologic Hashimoto's thyroiditis. *Mayo Clin. Proc.* 46:682–89
82. Morley, J. E., Jacobson, R. J., Melamed, J., Hershman, J. M. 1976. Choriocarcinoma as a cause of thyrotoxicosis. *Am. J. Med.* 60:1036–40
83. Kenimer, J. G., Hershman, J. M., Higgins, H. P. 1975. The thyrotropin in hydatidiform moles is human chorionic gonadotropin. *J. Clin. Endocrinol. Metab.* 40:482–91
84. McConnon, J. K., von Westarp, C., Mitchell, R. I. 1975. Follicular carcinoma of the thyroid with functioning metastases and hyperthyroidism. *Can. Med. Assoc. J.* 112:724–27
85. Beierwaltes, W. H. 1978. The treatment of thyroid carcinoma with radioactive iodine. *Semin. Nucl. Med.* 8:79–94
86. Blum, M., Kranjac, T., Park, C. M., Engleman, R. M. 1976. Thyroid storm after cardiac angiography with iodinated contrast medium. *J. Am. Med. Assoc.* 235:2324–25
87. Das, G., Krieger, M. 1969. Treatment of thyrotoxic storm with intravenous administration of propranolol. *Ann. Intern. Med.* 50:985–87
88. Williams, D. E., Chopra, I. J., Orgiazzi, J., Solomon, D. H. 1975. Acute effects of corticosteroids on thyroid activity in Graves' disease. *J. Clin. Endocrinol. Metab.* 41:354–61
89. Horn, K., Brehm, G., Huberman, J., Pickardt, C. R., Scriba, P. C. 1976. Successful treatment of thyrotoxic crisis by continuous plasmapheresis with blood cell separator. *Klin. Wochenschr.* 54: 983–86
90. Hermann, J., Kruskemper, H. L., Grosser, K. D., Hubner, W., Bohm, W. 1971. Peritoneal Dialyse in der Behandlung der thyreotoxischen Krise. *Deutsch. Med. Wochenschr.* 95:742–45
91. Larsen, P. R. 1972. Salicylate-induced increases in free triiodothyronine in man. *J. Clin. Invest.* 51:1125–34
92. Burrow, G. N. 1965. Neonatal goiter after maternal propylthiouracil therapy. *J. Clin. Endocrinol. Metab.* 25:403–8
93. Langer, A., Hung, C. T., McA'Nulty, J. A., Harrigan, J. T., Washington, E. 1974. Adrenergic blockade—a new approach to hyperthyroidism during pregnancy. *Obstet. Gynecol.* 44:181–86

Ann. Rev. Med. 1980. 31:405–19

BARTTER'S SYNDROME[1] ◆7369

John R. Gill, Jr., M.D.

Hypertension-Endocrine Branch, National Heart, Lung, and Blood Institute,
National Institutes of Health, Bethesda, Maryland 20014

In 1962, two patients with growth retardation, hypertrophy, and hyperplasia of the juxtaglomerular apparatus of the kidney, increased amounts of circulating angiotensin II, aldosteronism with hypokalemic alkalosis, and normal blood pressure with decreased pressor responsiveness to intravenous infusion of angiotensin II were reported as cases of a new syndrome that has since been called Bartter's syndrome (1). Since this description of the syndrome, the following additional abnormalities have been noted: pitressin-resistant impairment of urinary concentrating ability (1), decreased pressor responsiveness to intravenous infusion of norepinephrine (2), hypomagnesemia (3), hyperuricemia (4), and increased amounts of urinary prostaglandin E_2 (5).

The prevalence of Bartter's syndrome is not known, and, as only approximately 100 cases have been described over the last 18 years, the syndrome appears to be a rare cause of hypokalemia. The reported patients with Bartter's syndrome include blacks (1), Caucasians (6), and Orientals (7) and come from many geographic areas. Although it has been suggested that Bartter's syndrome occurs more frequently in blacks than in whites in the United States, where ten of fourteen reported families were black (8), such a preponderance of blacks has not been a feature of those patients with Bartter's syndrome studied at the National Institutes of Health, where only six of the eighteen families in this series were black. Thirteen of the eighteen patients in the NIH series were female, however, a preponderance that has been noted by others (9). The signs and symptoms of Bartter's syndrome have been noted as early as the first week of life (10) and as late as the fifth decade (11).

---

[1] The US Government has the right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper.

HISTOLOGICAL AND PATHOGENIC ASPECTS

The occurrence of Bartter's syndrome in siblings (3, 6, 10) and in offspring of consanguineous unions (3, 12, 13), with involvement of only a single generation, has led to the hypothesis that the syndrome is inherited as an autosomal recessive disorder. Siblings with the clinical features of Bartter's syndrome have also had histological changes such as severe glomerular and interstitial nephritis with mild juxtaglomerular hyperplasia and progressive renal failure (14) and normal glomeruli with normal appearance of the juxtaglomerular apparatus, interstitial fibrosis, and foam cells (15, 16) which suggest the presence of a nephritic process. These patients were thought to have a disorder other than Bartter's syndrome because juxtaglomerular hyperplasia was mild or absent (14, 16) and because of the presence of uremia, osteodystrophy and rapid course (14).

Histological features similar to those found in glomerulonephritis, such as enlarged hypercellular glomeruli, membranous thickening of glomerular capillary walls, mesangial hypercellularity, crescent formation, concentric periglomerular fibrosis, glomerulosclerosis and thickening, and sclerosis of arteries and arterioles have also been associated with juxtaglomerular hyperplasia in patients with Bartter's syndrome. These findings led some investigators to suggest that the syndrome may be acquired as a result of renal injury (9). Indeed, a number of lines of evidence support the view that renal injury, if it is not an etiologic factor in Bartter's syndrome, is at least a mediator of some of the renal histopathologic changes. First, a patient with the renal histological features of Bartter's syndrome at age six years (17), progressed to terminal renal failure at age 13 years and, at autopsy, diffuse glomerulosclerosis, proliferative glomerulitis, interstitial fibrosis, and sclerosis of renal arteries and arterioles were found; these changes were interpreted as chronic glomerulonephritis. In this regard, the findings of proliferative glomerulonephritis with electron-dense deposits in the basement membrane, which, on immunofluorescence, appeared to be immunoglobulins G and A and β_1C component of complement, have been observed in association with juxtaglomerular hyperplasia and Bartter's syndrome (18). Second, renal disease of known etiology, such as cystinosis, can present as Bartter's syndrome (19). Third, extrarenal disorders such as familial chloride diarrhea (20), surreptitious laxative abuse (21), and surreptitious vomiting (22) can present with all of the clinical features of Bartter's syndrome and the renal histopathologic changes of juxtaglomerular hyperplasia, periglomerular fibrosis, glomerulosclerosis, and thickened, hypercellular arterial and arteriolar walls with perivascular sclerosis. The presence in extrarenal disorders of renal morphological changes of Bartter's syndrome suggests that these changes may not be unique to Bartter's syndrome but represent a renal response to biochemical and physiological changes that the disorders have in common.

As the above discussion illustrates, Bartter's syndrome may have more than one etiology and may exhibit renal histological changes that are probably a composite of the underlying disease process and the renal response to biochemical and physiological changes. Presently available data suggest that a defect in chloride reabsorption in the thick ascending limb of the loop of Henle is the proximal pathophysiological event in Bartter's syndrome (23). The molecular abnormality which causes this defect in chloride transport is unknown; when the nature of the abnormality has been clarified, it may then be possible to resolve the questions about the role of heredity in the etiology of the syndrome and what renal histological changes, if any, are specific.

PATHOPHYSIOLOGY

Kidney and Adrenal

In the two index patients with Bartter's syndrome, a low sodium intake did not decrease urinary potassium excretion or increase serum potassium as it does in patients with primary aldosteronism (1). This failure of a restricted sodium intake to curtail potassium loss, a finding not only in those patients with a persistent renal sodium loss (11) but also in those who decreased their urinary sodium in a normal fashion (1, 12, 17), has been interpreted as evidence for an abnormality in the tubular handling of sodium. More specific information about the renal tubular reabsorption of sodium obtained by renal clearance studies during infusion of normal saline solution (24), hypotonic saline solution (11, 25–27), or dextrose and water (23, 28) indicated a decrease in proximal tubular reabsorption (11, 28), in distal tubular reabsorption (23–25, 27), or in both (23, 26, 28). The decrease in proximal tubular reabsorption observed in some patients during intravenous infusion of saline solution may be in part the result of changes in renal hemodynamics and in peritubular Starling forces (29, 30), and in part the result of impaired glomerular filtration (23). Whereas a decrease in proximal reabsorption was found in only some of the patients, a decrease in distal tubular reabsorption has been a consistent finding (23, 25–27). Data are not available to determine if those patients with Bartter's syndrome who cannot conserve sodium during a low sodium intake differ in any respect, other than in the degree of impairment of distal tubular reabsorption, from those who can conserve normally. One patient with obligatory sodium loss in whom tubular function has been assessed appeared to have a decrease in sodium reabsorption in the ascending limb of Henle's loop with normal proximal tubular function (25).

As most of the solute-free water is generated by active reabsorption of chloride in the ascending limb of the loop of Henle with sodium following passively (31, 32), measurements of free-water formation and of distal fractional chloride reabsorption provide an estimate of loop function. In patients with Bartter's syndrome, fractional solute-free water excretion and distal fractional chloride reabsorption were decreased, whereas in patients with a similar degree of hypokalemia and hyperreninemia as a result of gastrointestinal potassium loss, fractional free-water excretion and distal fractional chloride reabsorption were normal (23). (It should be noted, however, that an extreme degree of potassium depletion with serum potassium less than 2 meg/liter can produce a transient defect in chloride reabsorption (33).)

Sodium excretion can decrease in a normal fashion during a low sodium intake despite impaired reabsorption of sodium chloride in the loop of Henle. This suggests that other segments of the nephron can compensate for the defect to prevent abnormal losses of extracellular fluid (23, 24). Sodium excretion was abnormally high during water diuresis (23) and during infusion of saline (24), however, and this suggests that depression of sodium reabsorption by the proximal tubule, which occurs with expansion of extracellular fluid (29, 30), superimposed on impaired reabsorption of sodium chloride in the loop of Henle results in an exaggerated natriuresis that limits expansion of extracellular fluid.

Renal potassium wasting, a hallmark of Bartter's syndrome, is best explained in terms of the defect in tubular chloride reabsorption. Potassium is reabsorbed in the ascending limb of the loop of Henle and, presumably, its transport like that of sodium is passive and determined by the potential difference generated by active chloride transport (34). Also, an impairment in reabsorption in the loop of Henle, by leading to an increase in flow of tubular fluid at the potassium secretory site, can stimulate potassium secretion (34). If aldosterone production is increased, as it usually is in patients with Bartter's syndrome, this may further stimulate potassium secretion and contribute to potassium loss (34). Serum potassium remained low, however, when urinary aldosterone was normal (12, 27, 36), and bilateral adrenalectomy (10, 35) or treatment with aminoglutethimide, an inhibitor of aldosterone biogenesis (36), did not correct the hypokalemia. Thus aldosterone is an aggravating factor rather than the cause of renal potassium loss in Bartter's syndrome.

The impairment in chloride reabsorption in the loop of Henle may also be an important determinant of the decreased renal concentrating ability in Bartter's syndrome. Abstraction of water and concentration of tubular fluid in the collecting duct depends upon an intact reabsorptive process for sodium chloride in the loop of Henle to maintain medullary hypertonicity

(37). Thus, a defect in sodium chloride reabsorption in the loop of Henle with a loss of medullary interstitial hypertonicity can result in less abstraction of water out of the descending limb of Henle's loop as well as out of the collecting duct. Enhanced delivery of a larger, less-concentrated volume of tubular fluid to the ascending limb of Henle's loop and to more distal segments of the nephron may further limit sodium chloride reabsorption and the concentration of tubular fluid.

The increase in renal synthesis of prostaglandins, which occurs in patients with Bartter's syndrome (5, 38–40), could be a factor in the pathogenesis of the defect in sodium chloride reabsorption and in urinary concentrating ability by virtue of the ability of prostaglandin E to inhibit distal tubular reabsorption of sodium (42) and to oppose the action of vasopressin on the distal tubule (43). This does not appear to be the case, however, as treatment with an inhibitor of prostaglandin synthesis does not increase distal fractional chloride reabsorption (23) or maximal urinary osmolality (44).

Considerable evidence supports the hypothesis that prostaglandins mediate the hyperreninemia that characterizes Bartter's syndrome. Infusion of arachidonic acid, a precursor of prostaglandins, into the aorta of rabbits (45), and infusion of prostaglandins E_2 (46, 47) and I_2 (47) into a renal artery of the dog, increased the release of renin. Treatment of rabbit (45), dog (46, 48), and man (49) with inhibitors of prostaglandin synthesis decreased plasma renin activity. The addition of arachidonic acid (50), prostaglandin endoperoxides (50) and prostaglandin I_2 (51) to slices of rabbit renal cortex in vitro increased the release of renin into the incubation medium; addition of inhibitors of prostaglandin synthesis to the incubation medium inhibited the release of renin (50). These observations suggest that the decrease in plasma renin activity associated with the administration of indomethacin (5, 38, 39, 41), ibuprofen (5), and aspirin (40) to patients with Bartter's syndrome occurs as a result of inhibition of renal cortical prostaglandin synthesis. Although it may be argued that the retention of sodium and water, which also occurred in the patients with Bartter's syndrome during treatment with prostaglandin synthetase inhibitors, decreased plasma renin activity, this probably was not the case. Whereas it has been possible, on occasion, to decrease plasma renin activity to normal by means of a high sodium intake and rapid infusion of a large volume of saline (24), usually a high sodium intake or infusion of albumin produced only a partial decrease in plasma renin activity (11, 27, 36, 52), or none at all (7, 53, 54).

An increase in urinary prostaglandins is not unique to Bartter's syndrome but has been observed in other hypokalemic disorders such as psychogenic vomiting (22, 53, 55, 56) and experimental potassium depletion in the dog (48). The stimulatory effects of a low potassium concentration on renal prostaglandin synthesis in vitro (57, 58) as well as in vivo and the stimulatory effects of prostaglandins on renin release noted above suggest that prostaglandins may mediate the increase in plasma renin activity associated with potassium depletion (59, 60). The observation that treatment of potassium-depleted dogs with an inhibitor of prostaglandin synthesis decreased plasma renin activity without affecting the hypokalemia supports this conclusion (48).

In a similar fashion, treatment of patients with Bartter's syndrome with a prostaglandin synthetase inhibitor corrected the hyperreninemia and aldosteronism but not the hypokalemia (61) and provides evidence, in addition to that cited above (10, 35, 36), that aldosterone is not the primary cause of potassium loss in Bartter's syndrome. Although this discussion has centered on the role of potassium depletion in the pathogenesis of the hyperreninemia, it is also possible that increased delivery of sodium chloride to the macula densa, as a result of decreased chloride reabsorption in the ascending limb of the loop of Henle, also contributes to the hypersecretion of renin in patients with the syndrome (62).

Hypomagnesemia has been present in many but not all patients with Bartter's syndrome. Those affected have been young siblings (3) as well as older, sporadic patients and have not differed in any apparent way from those patients with a normal serum magnesium. Studies of the renal tubular handling of magnesium indicate that filtered magnesium is reabsorbed in proportion to sodium in the proximal tubule (63). Magnesium is also reabsorbed in the loop of Henle and it is this segment that regulates its rate of excretion into the urine (64). The process by which magnesium is transported out of the loop of Henle and the relation of its transport to that of chloride are not known. The association of a defect in chloride transport in the loop of Henle with urinary magnesium loss in patients with Bartter's syndrome suggests that the cause of the defect in chloride transport may also be capable of impairing magnesium transport in the loop of Henle.

Severe magnesium deficiency has been reported in association with hypokalemic alkalosis, hyperreninemia and high normal urinary aldosterone (65). Although it has been shown that experimental magnesium depletion in man leads to an increase in urinary potassium loss and a fall in serum potassium concentration (66), the cause of the potassium loss in magnesium depletion and the role of the renin-angiotensin-aldosterone systems are unknown. In any event, magnesium supplementation in hypomagnesemic patients with Bartter's syndrome appears to have little effect on the other manifestations of the disorder.

Hypercalciuria with a normal serum calcium has been observed in a few patients with Bartter's syndrome (54). The tubular reabsorption of calcium resembles that of magnesium in many respects (64), so that an abnormality in the loop of Henle may also impair the reabsorption of calcium in this segment of the nephron and increase its excretion in the urine.

The prevalence of hyperuricemia and acute gouty arthritis has been reported to be abnormally high in Bartter's syndrome (4). A high serum uric acid concentration has also been reported in patients with hypokalemic alkalosis and hyperreninemia caused by extrarenal losses of potassium and chloride (22, 55, 67), which suggests that biochemical and physiological changes may be the basis for the increase in serum uric acid. The observations that angiotensin II (68) and alkalosis (4) can each decrease the clearance of uric acid suggest that these factors may explain, at least in part, a decreased clearance of uric acid in Bartter's syndrome and other disorders characterized by hypokalemic alkalosis.

Blood Vessels and Blood

Normal blood pressure and resistance to the pressor effects of angiotensin II and, less consistently, norepinephrine are characteristic features of Bartter's syndrome. Although pressor resistance to angiotensin II was initially thought to be the proximal event in the pathogenesis of the syndrome (1), it has since been recognized that this finding is not unique to Bartter's syndrome but can occur in a variety of hypokalemic disorders (22, 48, 53, 55, 69). The observations that potassium depletion produces vasodilatation (48), and that plasma bradykinin, a vasodilator, is high in patients with Bartter's syndrome (70), suggest that increases in plasma renin activity and in adrenergic activity (71) are necessary to maintain blood pressure. A partial decrease in plasma renin activity in response to expansion of intravascular volume (11, 27, 36, 52), and a fall in blood pressure, at times marked, in response to intravenous infusion of an angiotensin II antagonist in patients with the syndrome (72, 73) support the notion that the increase in renin release may be, in part, a homeostatic response. Treatment of patients with Bartter's syndrome with a prostaglandin synthetase inhibitor corrects the hyperreninemia (5, 38–41), the increase in adrenergic activity (71), the increase in plasma bradykinin (70), the pressor resistance to angiotensin II (38, 39, 41, 53, 74), and to norepinephrine (53, 71, 74) without a change in blood pressure (5, 41).

Pressor responsiveness to exogenous angiotensin II is inversely related to plasma renin activity and to plasma angiotensin II; studies of this phenomenon suggest that a decrease in pressor responsiveness to angiotensin II is best explained by prior occupancy of vascular angiotensin receptors by endogenous angiotensin II rather than by a change in either the number of receptors or their affinity for angiotensin II (75). (It should be noted, however, that the phenomenon of down-regulation of receptor number cannot be excluded as a factor.) The increase in pressor responsiveness to angiotensin II during inhibition of prostaglandin synthesis in the patients with the syndrome was associated with a decrease in plasma renin activity. Thus it may be, in part, attributable to a decrease in occupancy of receptor sites by endogenous angiotensin II. It is also possible that the increase in pressor responsiveness to angiotensin II associated with inhibition of prostaglandin synthesis was, in part, the result of the restoration of plasma bradykinin to normal (70). Although the mechanism of the increase in plasma bradykinin is unknown (70), a direct relationship between plasma bradykinin and plasma renin activity has been observed in normal subjects and, presumably, modulates the pressor activity of angiotensin II (76). In addition to decreased receptor occupancy and decreased plasma bradykinin, inhibition of prostaglandin synthesis by blood vessels (77, 78), which may have been stimulated by angiotensin II (79), bradykinin (77, 79), and possibly by potassium depletion per se, could also have contributed to the increase in pressor responsiveness to angiotensin II, as prostaglandins synthesized by blood vessels oppose the vasoconstrictor effects of angiotensin II and norepinephrine (80). Thus, prior occupancy, supranormal plasma bradykinin, and increased synthesis of vasodilator prostaglandins by blood vessels may all be factors in the vascular resistance to angiotensin II in Bartter's syndrome. Increased synthesis of prostaglandins by blood vessels and supranormal plasma bradykinin are probably the major determinants of the vascular resistance to norepinephrine in Bartter's syndrome, as the occupancy by angiotensin II of its vascular receptors appears to have no effect on the vascular response to norepinephrine (81).

High values for urinary kallikrein (41, 70, 82) and low values for urinary kinins (70) have been found in Bartter's syndrome. Whereas the increase in kallikrein is probably, in part if not entirely, caused by aldosterone, the decrease in kinins appears to be mediated by prostaglandins and to be independent of urinary kallikrein (70). The relationship, if any, of urinary kallikrein and kinins to plasma bradykinin or to any other feature of the syndrome is unknown.

Increased sodium and decreased potassium concentrations have been found in the erythrocytes of some patients with Bartter's syndrome (10, 83). These changes in sodium concentration were associated with a decrease in fractional sodium outflux (83). As this defect in sodium transport, with and without an increase in erythrocyte sodium concentration, was also noted in otherwise normal parents and siblings of some of the patients with Bartter's syndrome, the defect was thought to be inherited and the cause of the increase in erythrocyte sodium concentration (83). In other studies of erythrocytes from two siblings with Bartter's syndrome, an increase in total ATPase activity with a low ratio of Na^+-K^+ ATPase to Mg^{2+} ATPase activity was observed; muscle cells from these two patients incubated in vitro showed a decreased potassium uptake (84). In related studies, sodium concentration in parotid saliva was increased in patients with Bartter's syndrome but decreased in those with primary aldosteronism and in those with hypokalemia, hyperreninemia, and aldosteronism produced by extrarenal potassium loss (85). Whether the impairment in sodium transport noted in a number of different cells in patients with Bartter's syndrome and in the erythrocytes of some of their unaffected relatives is a manifestation of the pathogenic process responsible for the decrease in chloride transport in the loop of Henle or is only a genetic marker for the syndrome (86) remains to be clarified.

Erythrocytosis and increased production of erythropoietin without polycythemia vera or apparent cause for secondary polycythemia has been reported in two patients with Bartter's syndrome (87, 88). The source of the erythropoietin was thought to be the hyperplastic juxtaglomerular apparatus but the reason for its increased release was unknown.

Studies reported in abstract indicate that platelets from patients with Bartter's syndrome show a decreased response to stimuli that induce aggregation in vitro (89, 90). As the defect in aggregation could be corrected by gel filtration of the platelets into normal plasma (90) or by treatment of the patients with aspirin (89), the defect was attributed to a prostaglandin or prostaglandin-like substance in the circulation.

Growth and Development (91)

When Bartter's syndrome occurs in infancy or childhood, it severely retards growth and may impair mental development. In those children who develop the syndrome during the first year of life, some degree of mental retardation has been a consistent finding. A delayed adolescent growth spurt usually occurs sometime after the seventeenth year with the eventual attainment of normal height. Hyperinsulinemia, with and without an abnormality in glucose tolerance, may be present. Treatment directed toward correction of the potassium deficiency appears to have little effect on growth and development or on the abnormalities of carbohydrate metabolism, although potassium depletion may well be the reason for the abnormalities.

DIFFERENTIAL DIAGNOSIS AND MANAGEMENT

Recent studies of three siblings indicate that renal potassium wasting with all of the clinical features of Bartter's syndrome can occur with the histological findings of normal glomeruli and juxtaglomerular apparati and with normal proximal and distal sodium chloride reabsorption (92). This disorder, apparently a familial tubulopathy but of unknown etiology, may be the same as that previously reported in two siblings (cases two and three in Reference 93), who showed minimal renal histological changes. It is not known, however, whether the potassium loss in these children was associated with a normal tubular reabsorption of sodium chloride as results of renal clearance studies were not reported. Determination of distal fractional chloride reabsorption (23) can be used to distinguish this disorder from Bartter's syndrome.

Vomiting, laxatives, or diuretics represent the most common causes of hypokalemia, and patients who vomit or take such medications surreptitiously may be indistinguishable from those with Bartter's syndrome (21, 22, 55). Although urinary potassium is usually low and points to the gastrointestinal tract as the source of the loss in patients with surreptitious vomiting or laxative abuse, urinary potassium can be inappropriately high if alkalosis or hypokalemia is severe, either because sodium bicarbonate is delivered to the distal tubule in large amounts or because the tubular handling of chloride is impaired by profound potassium depletion (33). In those patients with gastrointestinal loss of chloride and potassium and an inappropriately high urinary potassium, measurement of urinary chloride excretion during a chloride intake of approximately 100 meq per day can be a useful diagnostic test. If chloride excretion is low, usually less than 10 meq per day, and is associated with a much higher sodium and potassium excretion, a diagnosis of Bartter's syndrome is excluded and an abnormality of gastrointestinal function suggested (22, 55). Thus, urinary chloride determination can be very helpful and is usually successful as a means of distinguishing surreptitious vomiting or laxative abuse from Bartter's syndrome. The distinction between surreptitious diuretic abuse and Bartter's syndrome is more difficult. As the more commonly used diuretics such as furosemide and the thiazides decrease distal tubular reabsorption of sodium chloride, measurement of distal fractional chloride reabsorption in a patient who is taking them can show a decrease similar to that found in Bartter's syndrome. Therefore, a screen of the urine for the presence of diuretics may be necessary to confirm a suspicion of diuretic abuse or to establish a diagnosis of Bartter's syndrome.

Magnesium depletion, as noted above, can produce hypokalemia and needs to be considered in the differential diagnosis of Bartter's syndrome with hypomagnesemia (65, 66).

Familial chloride diarrhea is a rare disease that can present with the features of Bartter's syndrome (20, 67), but it can be readily distinguished by history, the character of the stools, and measurement of chloride in the urine (67).

The cornerstone of treatment of Bartter's syndrome is correction of potassium depletion and its associated symptoms. Oral administration of large amounts of potassium chloride (exceeding 160 meq per day of potassium) will usually correct the symptoms of weakness, tiredness, and easy fatigability although hypokalemia still persists. Such high dosages of potassium are usually not feasible because of gastric intolerance or diarrhea, and lower dosages have been given in conjunction with diuretics such as spironolactone or triamterene, which decrease the renal loss of potassium.

Hypomagnesemia, if present, should be treated with magnesium chloride. Although serum magnesium may show little change, symptoms of tetany and muscular cramps are frequently improved.

Propranolol decreased plasma renin activity and aldosterone production in some patients (94, 95) but not in others (39, 96), and it did not raise serum potassium even in those who showed a decrease in aldosterone (94, 95). Propranolol also usually produces a decrease in blood pressure, which at times may be of sufficient magnitude to produce symptoms, may fail to prevent the stimulation of renin release by spironolactone (95), and may produce a paradoxical increase in plasma renin activity and aldosterone production (39, 96). Therefore its use in Bartter's syndrome is probably not indicated.

Inhibitors of prostaglandin synthesis such as indomethacin (5, 38, 39, 41, 61, 70, 74), ibuprofen (5), or aspirin (40) can correct the hyperreninemia and aldosteronism in patients with Bartter's syndrome but may only partially correct the hypokalemia. As these inhibitors of prostaglandin synthesis may produce abdominal discomfort or peptic ulceration, marked fluid retention, and other distressing side effects, they must be used with care. Small dosages of prostaglandin synthetase inhibitors combined with supplemental potassium and potassium-sparing diuretics may permit a greater degree of correction of potassium deficiency in association with a lower plasma renin activity than has been heretofore possible (97). Such a therapeutic approach can stimulate growth and development in children with Bartter's syndrome (97) and may retard the development of proliferative and sclerotic changes in renal arteries, arterioles and glomeruli and renal failure which can occur in the syndrome.

Patients with Bartter's syndrome can become pregnant (98, 99), and the disorder may either improve spontaneously during pregnancy (98) or be readily managed (99) without outward effects on the fetus.

Literature Cited

1. Bartter, F. C., Pronove, P., Gill, J. R. Jr., MacCardle, R. C. 1962. Hyperplasia of the juxtaglomerular complex with hyperaldosteronism and hypokalemic alkalosis. *Am. J. Med.* 33:811–28
2. Gill, J. R. Jr., Bartter, F. C. 1964. In *Water and Electrolyte Metabolism (II)*, ed. J. deGraeff, B. Leijnse, pp. 119–24. Amsterdam: Elsevier. 251 pp.
3. Sutherland, L. E., Hartroft, P., Balis, J. U., Bailey, J. D., Lynch, M. J. 1970. Bartter's syndrome. *Acta Paediatr. Scand.* Suppl 201, pp. 1–24
4. Meyer, W. J. III, Gill, J. R. Jr., Bartter, F. C. 1975. Gout as a complication of Bartter's syndrome. *Ann. Intern. Med.* 83:56–59
5. Gill, J. R. Jr., Frölich, J. C., Bowden, R. E., Taylor, A. A., Keiser, H. R., Seyberth, H. W., Oates, J. A., Bartter, F. C. 1976. Bartter's syndrome: A disorder characterized by high urinary prostaglandins and a dependence of hyperreninemia on prostaglandin synthesis. *Am. J. Med.* 61:43–51

6. Visser, H. K. A., Degenhart, H. J., Desmit, E., Cost, W. S. 1967. Mineralocorticoid excess in two brothers with dwarfism, hypokalemic alkalosis and normal blood pressure. *Acta Endocrinol. Copenhagen* 55:661–72
7. Imai, M., Yabuta, K., Murata, H., Takita, S., Ohbe, Y., Sokabe, H. 1969. A case of Bartter's syndrome with abnormal renin response to salt load. *J. Pediatr.* 74:738–49
8. Hall, B. D. 1971. Preponderance of Bartter's syndrome among blacks. *N. Engl. J. Med.* 285:581
9. Cannon, P. J., Leeming, J. M., Sommers, S. C., Winters, R. W., Laragh, J. H. 1968. Juxtaglomerular hyperplasia and secondary hyperaldosteronism (Bartter's syndrome): A reevaluation of the pathophysiology. *Medicine* 47: 107–31
10. Trygstad, C. W., Mangos, J. A., Bloodworth, J. M. B., Lobeck, C. C. 1969. A sibship with Bartter's syndrome: failure of total adrenalectomy to correct the potassium wasting. *Pediatrics* 44: 234–42
11. Tomko, D. J., Yeh, B. P. Y., Falls, W. F. Jr. 1976. Bartter's syndrome. Study of a 52 year old man with evidence for a defect in proximal tubular sodium reabsorption and comments on therapy. *Am. J. Med.* 61:111–17
12. VanOlphen, A. H. F., VanMunster, P. J. J., Slooff, J. P. 1964. See Ref. 2, pp. 134–47
13. Sotos, J. F. 1970. In *Genetic Disorders of Man*, ed. R. M. Goodman, pp. 739–41. Boston: Little, Brown. 1009 pp.
14. Arant, B. S., Brackett, N. C. Jr., Young, R. B., Still, W. J. S. 1970. Case studies of siblings with juxtaglomerular hyperplasia and secondary aldosteronism associated with severe azotemia and renal rickets—Bartter's syndrome or disease? *Pediatrics* 46:344–61
15. France, R., Merrill, J. M., Norris, J. L., Tolleson, W. J. 1959. Further studies in a case of potassium depletion of unknown cause. *Trans. Am. Clin. Climatol. Assoc.* 71:45–51
16. Potter, W. Z., Trygstad, C. W., Helmer, O. M., Nance, W. E., Judson, W. E. 1974. Familial hypokalemia associated with renal interstitial fibrosis. *Am. J. Med.* 57:971–77
17. Bryan, G. T., MacCardle, R. C., Bartter, F. C. 1966. Hyperaldosteronism, hyperplasia of the juxtaglomerular complex, normal blood pressure and dwarfism: Report of a case. *Pediatrics* 37:43–50
18. LaGreca, G., Bazzato, G., Biasioli, S., Coli, U., Fabris, A., Gaianigo, P., Mantero, F., Olivetti, G., Pisani, E., Bonanno, A. 1975. Bartter's syndrome associated with sponge disease. *Kidney Int.* 73:192
19. Berio, A. 1978. Considerazioni sulla nefropatio della cistinosi con s. di Bartter Secondaria. *Minerva Pediatr.* 30: 1825–31
20. Pasternack, A., Perheentupa, J., Launiala, K., Hallman, N. 1967. Kidney biopsy findings in familial chloride diarrhoea. *Acta Endocrinol.* 55:1–9
21. Fleischer, N., Brown, H., Graham, D. Y., Delena, S. 1969. Chronic laxative-induced hyperaldosteronism and hypokalemia simulating Bartter's syndrome. *Ann. Intern. Med.* 70:791–98
22. Ramos, E., Hall-Craggs, M., Demers, L. M. 1979. A case of surreptitious, habitual vomiting simulating Bartter's syndrome. *J. Am. Med. Assoc.* In press
23. Gill, J. R. Jr., Bartter, F. C. 1978. Evidence for a prostaglandin-independent defect in chloride reabsorption in the loop of Henle as a proximal cause of Bartter's syndrome. *Am. J. Med.* 65: 766–72
24. White, M. G. 1972. Bartter's syndrome. A manifestation of renal tubular defects. *Arch. Intern. Med.* 129:41–47
25. Chaimovitz, C., Levi, J., Bertter, O. S., Oslander, L., Benderli, A. 1973. Studies on the site of renal salt loss in a patient with Bartter's syndrome. *Pediatr. Res.* 7:89–94
26. Chan, J. C. M., Malekzedeh, M. H., Anand, S. K. 1975. Defect in renal tubular sodium reabsorption in a patient with Bartter's syndrome. *Clin. Proc. Child. Hosp. Natl. Med. Cent.* 31:67–71
27. Fujita, T., Sakuguchi, H., Shibagaki, M., Fukui, T., Nomura, M., Sekiguchi, S. 1977. The pathogenesis of Bartter's syndrome. *Am. J. Med.* 63:467–74
28. Bartter, F. C., Delea, C. S., Kowasaki, T., Gill, J. R. Jr. 1974. The adrenal cortex and the kidney. *Kidney Int.* 6:272–80
29. Martino, J. A., Earley, L. E. 1967. Demonstration of a role of physical factors as determinants of the natriuretic response to volume expansion. *J. Clin. Invest.* 1963–78
30. Brenner, B. M., Falchuk, K. H., Keimowitz, R. I., Berliner, R. W. 1969. The relationship between peritubular capillary protein concentration and fluid reabsorption by the renal proximal tubule. *J. Clin. Invest.* 48:1519–31

31. Burg, M., Green, N. 1973. Function of the thick ascending limb of Henle's loop. *Am. J. Physiol.* 224:659–68
32. Rocha, A. S., Kokko, J. P. 1973. Sodium chloride transport in the medullary thick ascending limb of Henle. *J. Clin. Invest.* 52:612–23
33. Garella, S., Chazan, J. A., Cohen, J. J. 1970. Saline-resistant metabolic alkalosis or "chloride-wasting nephropathy." *Ann. Intern. Med.* 73:31–38
34. Wright, F. S., Giebisch, G. 1978. Renal potassium transport: contributions of individual nephron segments and populations. *Am. J. Physiol.* 235:F515–F527
35. Takayasu, H., Aso, Y., Lakanchi, K. 1971. A case of Bartter's syndrome with surgical treatment followed for four years. *J. Clin. Endocrinol. Metab.* 32: 842–45
36. Goodman, A. D., Vagnucci, A. H., Hartroft, P. M. 1969. Pathogenesis of Bartter's syndrome. *N. Engl. J. Med.* 281:1435–39
37. Jamison, R. L. 1976. In *The Kidney,* ed. B. M. Brenner, F. C. Rector Jr., pp. 391–441. Philadelphia: Saunders. 1948 pp.
38. Verberckmoes, R., VanDamme, B., Clement, J., Avery, A., Michielsen, P. 1976. Bartter's syndrome with hyperplasia of renomedullary cells: successful treatment with indomethacin. *Kidney Int.* 9:302–7
39. Fichman, M. P., Telfer, N., Zia, P., Speckart, P., Golub, M., Rude, R. 1976. Role of prostaglandins in the pathogenesis of Bartter's syndrome. *Am. J. Med.* 60:785–97
40. Norby, L., Flamenbaum, W., Lentz, R., Ramwell, P. 1976. Prostaglandins and aspirin therapy in Bartter's syndrome. *Lancet* 2:604–6
41. Halushka, P. V., Wohltmann, H., Privitera, P. J., Hurwitz, G., Margolius, H. S. 1977. Bartter's syndrome: Urinary prostaglandin E-like material and kallikrein, indomethacin effects. *Ann. Intern. Med.* 87:281–86
42. Stokes, J. B., Kokko, J. P. 1977. Inhibition of sodium transport by prostaglandin E_2 across the isolated, perfused rabbit collecting tubule. *J. Clin. Invest.* 59:1099–1104
43. Lum, G. M., Aisenbrey, G. A., Dunn, M. J., Berl, T., Schrier, R. W., McDonald, K. M. 1977. *In vivo* effect of indomethacin to potentiate the renal medullary cyclic AMP response to vasopressin. *J. Clin. Invest.* 59:8–13
44. Chan, J. C. M., Gill, J. R. Jr., Bartter, F. C. 1979. Bartter's syndrome: prostaglandin inhibition and renal concentration. *Pediatr. Res.* 13:511
45. Larsson, C., Weber, P., Änggård, E. 1974. Arachidonic acid increases and indomethacin decreases plasma renin activity in the rabbit. *Eur. J. Pharmacol.* 28:391–94
46. Yun, J., Kelly, G., Bartter, F. C., Smith, H. Jr. 1977. Role of prostaglandins in the control of renin release. *Circ. Res.* 40:459–64
47. Gerber, J. G., Branch, R. A., Nies, A. S., Gerkens, J. F., Shand, D. G., Hollifield, J., Oates, J. A. 1978. Prostaglandins and renin release: II. Assessment of renin secretion following infusion of PGI_2, E_2, and D_2 into the renal artery of anesthetized dogs. *Prostaglandins* 15: 81–88
48. Galvez, O. G., Bay, W. H., Roberts, B. W., Ferris, T. F. 1977. The hemodynamic effects of potassium deficiency in the dog. *Circ. Res.* 40:Suppl I: I-11–I-16
49. Frölich, J. C., Hollifield, J. W., Dormois, J. C., Frolich, B. L., Seyberth, H., Michelakis, A. M., Oates, J. A. 1976. Suppression of plasma renin activity by indomethacin in man. *Circ. Res.* 39: 447–52
50. Weber, P. C., Larsson, C., Änggård, E., Hamberg, M., Covey, E. J., Nicolaou, K. C., Samuelsson, B. 1976. Stimulation of renin release from rabbit renal cortex by arachidonic acid and prostaglandin endoperoxides. *Circ. Res.* 39:868–74
51. Whorton, A. R., Misono, K., Hollifield, J., Frölich, J. C., Inagami, T., Oates, J. A. 1977. Prostaglandins and renin release: I. Stimulation of renin release from rabbit renal cortical slices by PGI_2. *Prostaglandins* 14:1095–1104
52. Brackett, N. C. Jr., Koppel, M., Randall, R. E. Jr., Nixon, W. P. 1968. Hyperplasia of the juxtaglomerular complex with secondary aldosteronism without hypertension (Bartter's syndrome). *Am. J. Med.* 44:803–19
53. Radfar, N., Gill, J. R. Jr., Bartter, F. C., Bravo, E., Taylor, A. A., Bowden, R. E. 1978. Hypokalemia, in Bartter's syndrome and other disorders, produces resistance to vasopressors via prostaglandin overproduction. *Proc. Soc. Exp. Biol. Med.* 158:502–7
54. McCredie, D. A., Rotenberg, E., Williams, A. C. 1974. Hypercalciuria in potassium-losing nephropathy: A variant of Bartter's syndrome. *Aust. Paediatr. J.* 10:286–95
55. Veldhuis, J. D., Bardin, C. W., Demers, L. M. 1979. Metabolic mimicry of Bart-

ter's syndrome by covert vomiting. *Am. J. Med.* 66: 361–363

56. Nivet, H., Gremier, B., Rolland, J. C., Lebranchu, Y. 1978. Raised urinary prostaglandins in a patient without Bartter's syndrome. *Lancet* 1:333–34
57. Zusman, R. M., Keiser, H. R. 1977. Prostaglandin biosynthesis by rabbit renomedullary interstitial cells in tissue culture. Stimulation by angiotensin II, bradykinin and arginine vasopressin. *J. Clin. Invest.* 60:215–23
58. Düsing, R., Attallah, A. A., Prezyna, A. P., Lee, J. B. 1978. Renal biosynthesis of prostaglandin E_2 and $F_{2\alpha}$: Dependence on extracellular potassium. *J. Lab. Clin. Med.* 92:669–77
59. Sealey, J. E., Clark, I., Bull, M. R., Laragh, J. H. 1970. Potassium balance and the control of renin secretion. *J. Clin. Invest.* 49:2119–27
60. Brunner, H. R., Baer, L., Sealey, J. E., Ledingham, J. G. G., Laragh, J. H. 1970. The influence of potassium administration and of potassium deprivation on plasma renin in normal and hypertensive subjects. *J. Clin. Invest.* 49:2128–38
61. Bowden, R. E., Gill, J. R. Jr., Radfar, N., Taylor, A. A., Keiser, H. R. 1978. Prostaglandin synthetase inhibitors in Bartter's syndrome. *J. Am. Med. Assoc.* 239:117–21
62. Schnermann, J., Ploth, D. W., Hermle, M. 1976. Activation of tubulo-glomerular feedback by chloride transport. *Pflugers Arch.* 363:229–40
63. Brunette, M., Wen, S.-F., Evanson, R. L., Dirks, J. H. 1969. Micropuncture study of magnesium reabsorption in the proximal tubule of the dog. *Am. J. Physiol.* 216:1510–16
64. Murayama, Y., Morel, F., LeGrimellec, C. 1972. Phosphate, calcium and magnesium transfers in proximal tubules and loops of Henle as measured by single nephron microperfusion experiments in the rat. *Pflugers Arch.* 333:1–16
65. Gitelman, H. J., Graham, J. B., Welt, L. G. 1966. A new familial disorder characterized by hypokalemia and hypomagnesemia. *Trans. Assoc. Am. Phys.* 79:221–23
66. Shils, M. E. 1969. Experimental human magnesium depletion. *Medicine* 48:61–85
67. Pearson, A. J. G., Sladen, G. E., Edmonds, C. J., Tavill, A. S., Wills, M. R., McIntyre, N. 1973. The pathophysiology of congenital chloridorrhoea. *Q. J. Med.* 167:453–66
68. Ferris, T. F., Gorden, P. 1968. Effect of angiotensin and norepinephrine upon urate clearance in man. *Am. J. Med.* 44:359–65
69. Burmeister, H., Ruiz-Torres, A. 1971. Zur frage der angiotensin resistenz: Kritische anmerkung zum sogenannten Bartter-syndrom. *Klin. Wochenschr.* 49:1275–79
70. Vinci, J. M., Gill, J. R. Jr., Bowden, R. E., Pisano, J. J., Izzo, J. L. Jr., Radfar, N., Taylor, A. A., Zusman, R. M., Bartter, F. C., Keiser, H. R. 1978. The kallikrein-kinin system in Bartter's syndrome and its response to prostaglandin synthetase inhibition. *J. Clin. Invest.* 51:1671–82
71. Silverberg, A. B., Mennee, P. A., Cryer, P. E. 1978. Resistance to endogenous norepinephrine in Bartter's syndrome. *Am. J. Med.* 64:231–35
72. Kono, T., Oseko, F., Shimbo, S., Nanno, M., Ikeda, F., Endo, J. 1976. Blood pressure fall by angiotensin II antagonist in patients with Bartter's syndrome. *J. Clin. Endocrinol.* 43:692–95
73. Sasaki, H., Okumura, M., Asano, T., Arakawa, K., Kawasaki, T. 1977. Responses to angiotensin II antagonist before and after treatment with indomethacin in Bartter's syndrome. *Br. Med. J.* 2:975–1038
74. Bartter, F. C., Gill, J. R. Jr., Frölich, J. C., Bowden, R. E., Hollifield, J. W., Radfar, N., Keiser, H. R., Oates, J. A., Seyberth, H., Taylor, A. A. 1976. Prostaglandins are overproduced by the kidneys and mediate hyperreninemia in Bartter's syndrome. *Trans. Assoc. Am. Phys.* 89:77–90
75. Thurston, H. 1976. Vascular angiotensin receptors and their role in blood pressure control. *Am. J. Med.* 61: 768–78
76. Wong, P. Y., Talamo, R. C., Williams, G. H., Coleman, R. W. 1975. Response of kallikrein-kinin and renin-angiotensin systems to saline infusion and upright posture. *J. Clin. Invest.* 55:691–98
77. Terragno, D. A., Crowshaw, K., Terragno, N. A., McGiff, J. C. 1975. Prostaglandin synthesis by bovine mesenteric arteries and veins. *Circ. Res.* 36, 37: Suppl 1, I-76–I-80
78. Gimbrone, M. A. Jr., Alexander, R. W. 1975. Angiotensin II stimulation of prostaglandin production in cultured human vascular endothelium. *Science* 189:219–20
79. Alexander, R. W., Gimbrone, M. A. Jr. 1976. Stimulation of prostaglandin E synthesis in cultured human umbilical

vein smooth muscle cells. *Proc. Natl. Acad. Sci. USA* 73:1617–20

80. Vane, J. R., McGiff, J. C. 1975. Possible contributions of endogenous prostaglandins to the control of blood pressure. *Circ. Res.* 36, 37: Suppl 1, I-68–I-75
81. Oliver, J. A., Cannon, P. J. 1978. The effect of altered sodium balance upon renal vascular reactivity to angiotensin II and norepinephrine in the dog. *J. Clin. Invest.* 61:610–23
82. Lechi, A., Covi, G., Lechi, C., Mantero, F., Scuro, L. A. 1976. Urinary kallikrein excretion in Bartter's syndrome. *J. Clin. Endocrinol. Metab.* 43:1175–78
83. Gardner, J. D., Simopulos, A. P., Lapey, A., Shibolet, S. 1972. Altered membrane sodium transport in Bartter's syndrome. *J. Clin. Invest.* 51: 1565–71
84. Haljamae, H., Enger, E., Sigstrom, L. 1975. Cellular potassium transport and ATPase activity in Bartter's syndrome. *Scand. J. Clin. Lab. Invest.* 35:53–58
85. Heidland, A., Kreusser, W., Hennemann, H., Knauf, H., Wigand, M. E. 1972. Excretion of the mono- and divalent ions in relation to flow in Bartter's and pseudo-Bartter's syndrome in parotid saliva. Comparative studies to the syndrome of Conn. *Klin. Wochenschr.* 50:959–66
86. Bartter, F. C. 1974. The syndrome of juxtaglomerular hyperplasia with aldosteronism, hypokalemic alkalosis and normal blood pressure. *Birth Defects* 10:104–8
87. Jepson, J., McGarry, E. E. 1968. Polycythemia and increased erythropoietin production in a patient with Bartter's syndrome. *Blood* 32:370–75
88. Erkelens, D. W., Statius vanEps, L. W. 1973. Bartter's syndrome and erythrocytosis. *Am. J. Med.* 55:711–19
89. O'Regan, S., Rivard, G. E., Robitaille, P. O. 1978. Aspirin improves platelet dysfunction in Bartter's syndrome. *Abstr. 7th Int. Congr. Nephrol.*, p. 1
90. MacIntyre, D., Smith, M., Levine, L., Stoff, J., Stemerman, M., Steer, M., Salzman, E., Brown, R. 1979. Is the platelet defect in Bartter's syndrome associated with a plasma prostaglandin? *Abstr. 4th Int. Prostaglandin Conf.*, p. 72
91. Simopoulos, A. P. 1979. Growth characteristics in patients with Bartter's syndrome. *Nephron* 23:131–36
92. Güllner, H.-G., Gill, J. R. Jr., Bartter, F. C., Chan, J. C. M., Dickman, P. S. 1979. A familial disorder with hypokalemic alkalosis, hyperreninemia, aldosteronism, high urinary prostaglandins and normal blood pressure that is not "Bartter's syndrome." *Trans. Assoc. Am. Phys.* 92:175–88
93. McCredie, D. A., Blair-West, J. R., Scoggins, B. A., Shipman, R. 1971. Potassium-losing nephropathy of childhood. *Med. J. Aust.* 1:129–35
94. Modlinger, R. S., Nicolis, G. L., Krakoff, L. R., Gabrilove, J. L. 1973. Some observations on the pathogenesis of Bartter's syndrome. *N. Engl. J. Med.* 289:1022–24
95. Solomon, R. J., Brown, R. S. 1975. Bartter's syndrome. *Am. J. Med.* 59: 575–83
96. Schwartz, G. J., Cornfeld, D. 1974. Propranolol in Bartter's syndrome. *N. Engl. J. Med.* 290:966
97. Littlewood, J. M., Lee, M. R., Meadow, S. R. 1978. Treatment of Bartter's syndrome in early childhood with prostaglandin synthetase inhibitors. *Arch. Dis. Child.* 53:43–48
98. Klaus, D., Klumpp, F., Rossler, R. 1971. Einfluss der schwangerschaft auf das Bartter-syndrom. *Klin. Wochenschr.* 49:1280–85
99. Stokes, G. S., Andrews, B. S., Hagon, E., Thornell, I. R., Palmer, A. A., Posen, S. 1974. Bartter's syndrome presenting during pregnancy. *Med. J. Aust.* 2:360–65

Ann. Rev. Med. 1980. 31:421–31
Copyright © 1980 by Annual Reviews Inc. All rights reserved

PRIMARY PULMONARY HYPERTENSION[1]

❖7370

Alfred P. Fishman, M.D. and Giuseppe G. Pietra, M.D.

Cardiovascular-Pulmonary Division, Departments of Medicine
and Pathology, University of Pennsylvania School of Medicine,
Philadelphia, Pennsylvania 19104

INTRODUCTION

The designation "primary pulmonary hypertension" is generally used as a synonym for "unexplained pulmonary hypertension" (1, 2). Both clinicians and pathologists have adopted this term. Unfortunately, despite this apparent concurrence, not all who use the term have in mind the same disorders, etiologies, pathogenetic mechanisms, or anatomic lesions.

For the clinician, the syndrome of primary pulmonary hypertension has three essential features: (*a*) clinical, x-ray, and electrocardiographic evidence of pulmonary hypertension, (*b*) abnormally high pulmonary arterial pressures and pulmonary vascular resistance in the face of a normal pulmonary wedge pressure and a nearly normal cardiac output, and (*c*) insufficient cause for the pulmonary hypertension either in the lungs or heart (3).

Since the advent of cardiac catheterization (4), the pathologist has participated in the search for etiology by seeking to establish the extent and nature of the widespread pulmonary vascular disease, and by excluding cardiac or pulmonary causes of diffuse pulmonary vascular involvement. Indeed, recently a few pathologists assumed the lead in attempting to sort out primary pulmonary hypertension when they proposed that a distinctive vascular lesion, i.e. "plexogenic arteriopathy," is prerequisite for the diagnosis of primary pulmonary hypertension (1). But, as indicated below, this proposition has not been universally accepted (5–7).

---

[1] The material in this review was presented in part in the Dickinson W. Richards Memorial Lecture, at the Annual Meeting of the American Heart Association, 1979.

0066-4219/80/0401-0421$01.00

EPIDEMIOLOGY

About 1000 cases have been reported to date (8). Except for the single epidemic associated with the ingestion of the anorexigenic agent, Aminorex®, the disorder has been sporadic. In busy cardiac catheterization laboratories devoted primarily to right heart catheterization, the incidence of primary pulmonary hypertension is of the order of 1–2% (9–11) and is apparently strongly influenced by the population that a particular laboratory encounters.

Because pulmonary hypertension is apt to be overlooked clinically unless it is severe, the age of onset is often difficult to define accurately. Until adolescence, males and females are about equally affected; thereafter, females predominate (10, 12). The age span is exceedingly broad, ranging from a few days after birth to the sixth decade (8). Most published reports deal with adults in their twenties or thirties. The wide spread in age may reflect different etiologies. This view is strongly supported by the spikes in incidence in young children and women and by the epidemic associated with the ingestion of aminorex (9).

NATURAL HISTORY

Except for its late stage, the course of primary pulmonary hypertension is difficult to trace because (*a*) it is an uncommon disorder, (*b*) it is usually discovered late in its evolution, and (*c*) the course may vary with the etiologic agent. For these reasons, unless one deals with a specific group of people with primary pulmonary hypertension (9, 13), survival is not easily related to the patient's age at onset. For example, in one series the mean survival time was two to three years whereas in others it was 19–29 years (12, 14). Nor is the usual grim prediction warranted that, once diagnosed, rapid progression to death is inevitable. Indeed, in children (15) and in adults after aminorex ingestion (9, 10), reversal is not uncommon.

ETIOLOGY

By definition, the clinical designation primary pulmonary hypertension assumes that despite exhaustive search, the cause(s) is still a mystery and that the disease is confined to the arterial resistance vessels. Occasionally, lung biopsy or autopsy solves the mystery by disclosing subtle, diffuse pulmonary interstitial fibrosis, inapparent on the chest x ray, as the cause of the pulmonary hypertension (16). At other times, occlusive pulmonary vascular disease proves to be part of a systemic disorder, particularly collagen vascular disease (8). Also, an occasional instance of progressive systemic sclerosis causes malignant pulmonary hypertension even though intimal fibrosis is minimal or absent (17). Sometimes the enigma of primary pulmonary hypertension is compounded by the coexistence of systemic and pulmonary hypertension (18), which raises the prospect that a similar humoral or autonomic nervous mechanism may be involved in both. In tropical countries, obliterative vascular disease caused by the passage of blood-borne parasites through the lungs sometimes evokes a syndrome that is indistinguishable from primary pulmonary hypertension (19). This occurrence reinforces the widely held suspicion that many vascular insults may culminate in the same types of obliterative pulmonary vascular disease.

In some patients, such as those in whom primary pulmonary hypertension is associated with Raynaud's phenomenon, the pulmonary hypertension appears to be part of a generalized vascular disorder (20, 21). But, even when all known causes of secondary hypertension, no matter how subtle (22), have been excluded on clinical grounds and by lung biopsy, primary pulmonary hypertension remains a heterogeneous group of disorders. Three particular aspects of this syndrome suggest that different mechanisms may be operative in individual patients: (*a*) the familial incidence, (*b*) the different ages of onset, and (*c*) outbursts associated with the ingestion of particular drugs or foods.

Familial Incidence

Primary pulmonary hypertension sometimes runs in families (8, 23, 24); it also occurs in twins (25). Whether the disorder is inherited (26, 27) or acquired from some shared environmental exposure is not clear.

The Age of Onset

One variant of the syndrome occurs soon after birth. It has been widely attributed to persistence of the fetal circulation (28, 29). Other mechanisms under consideration are fetal systemic hypertension or constriction of the ductus arteriosus in utero (28). The outstanding histological feature is hypertrophy of the media in small pulmonary arteries in a pattern reminiscent of that seen in response to chronic hypoxia (30); fibrous obliteration of pulmonary vessels is inconspicuous or absent (31, 32). Plexiform lesions have not been described. The disorder presumably reflects either failure of hypertrophied fetal vessels to involute after birth or a disturbance in prostaglandin metabolism. The primary pulmonary hypertension of early childhood has no known relationship to primary pulmonary hypertension that affects young or middle-aged women. Nor is either of these clearly related to primary pulmonary hypertension in young men.

Drug- or Food-Associated Mechanisms

Probably the strongest clue that primary pulmonary hypertension can have multiple etiologies and that our way of life may contribute arcanely to its pathogenesis was provided by an outbreak of pulmonary hypertension in Switzerland, Austria, and Germany between 1966 and 1968 (9–11, 33). The epidemic followed the introduction in November 1965 of an appetite depressant agent, aminorex (5-amino-phenyloxazoline); the epidemic subsided after aminorex was withdrawn from the market in 1968.

Aminorex resembles epinephrine and amphetamine in chemical structure. Attempts to reproduce the disorder in experimental animals by administering aminorex or catecholamines have been consistently unsuccessful (34). Despite some uncertainty, aminorex is generally accepted as the cause of the epidemic and as proof that an agent taken by mouth can selectively affect the pulmonary circulation. One residual enigma is why so few of the obese individuals who took the drug developed pulmonary hypertension; differences in individual susceptibility—presumably genetic—have been postulated (9). And to explain why the disease has not been reproducible experimentally, a puzzling species difference has been invoked (35).

The pulmonary vascular lesions associated with aminorex ingestion satisfy all anatomical criteria for primary pulmonary hypertension. Arterioles, from 40 to 600μm (3, 8, 11, 36) are predominantly affected. The lesions run the full range from medial hypertrophy and intimal fibrosis including "onion skin" lesions on the one hand, to focal necrosis of the vessel wall, perivascular inflammation, and plexiform lesions on the other. As expected, secondary phenomena of hyalinization of arteriolar walls, fibrin deposition, organization, and recanalization are superimposed. Vasoconstriction has been inferred on morphologic, as well as physiologic, grounds to contribute to the pulmonary hypertension (8). The large pulmonary arteries show structural changes consistent with the high pressures to which they are exposed (8).

The aminorex experience pinpointed the possibility that substances taken by mouth could cause pulmonary hypertension that would defy attempts at conventional explanation (37). In support of this notion is the pulmonary hypertension produced by feeding the seeds of the shrub, *Crotalaria spectabilis,* to the rat, dog, and monkey (34, 38–40). The offending agents in the various types of crotalaria are the pyrrolizidine alkaloids: in *C. spectabilis,* the alkaloid is monocrotaline; in *C. fulva,* it is fulvine.

The largest experimental experience has been with monocrotaline (38). This alkaloid does not act directly on the pulmonary circulation; instead it seems to be converted by the liver to dehydromonocrotaline before exerting its pulmonary toxicity. At autopsy, widespread interstitial fibrosis and medial hypertrophy of the small pulmonary arteries coexists with perivascular fibrosis; plexiform lesions and intimal fibrosis are absent. These findings suggest that the pulmonary hypertension is secondary to interstitial fibrosis and vasculitis (39).

Other evidence for "dietary" pulmonary hypertension is sparse. Bush tea has been implicated in the pathogenesis of pulmonary veno-occlusive disease in both animals and man (38). Phenformin has also been indicated as an occasional cause (41). Despite this rarity, the experience with aminorex and bush tea on the one hand and crotalaria on the other suggest that substances taken by mouth can cause obliterative vascular lesions in the pulmonary circulation. How this gastrointestinal-liver-lung axis operates is unclear. Is it a metabolite of ingested food? Or are the ingested materials disrupting the processing of circulating vasoactive substances, such as serotonin, prostaglandins, or nucleotides, by the liver or the lungs?

Of great interest along this line is the coincidence of portal and pulmonary hypertension (42, 43). Although some instances are undoubtedly due to, or complicated by, thromboembolic disease from the shunt (44), in others the pulmonary hypertension seems primary in the lungs. This coincidence, as well as the association of aminorex and pulmonary hypertension, again raises the question of individual predisposition since so few patients with severe liver disease develop pulmonary hypertension. Complicating the interpretation of an association between pulmonary and portal hypertension is the observation that acute hypoxia fails to elicit pulmonary vasoconstriction in patients with hepatic cirrhosis (45).

If the syndrome of dietary pulmonary hypertension does exist, how does it come about? Multiple thromboemboli cannot be invoked, since the vascular lesions do not appear to be thromboembolic in origin. Perhaps subtle endothelial injury and failure to inactivate vasoconstrictor substances brought to the lungs is at fault. After the original insult to the endothelium, is vascular damage self-perpetuating because of platelet aggregation, local release of vasoactive substances, endothelial dehiscence, naked basement membranes, and obstruction of the minute pulmonary vessels? Can endothelial injury be the common denominator by which certain drugs, such as nitrofurantoin, busulfan, monocrotaline, or hydralazine stimulate interstitial fibrosis in predisposed individuals?

PATHOLOGY

For pulmonary artery pressures to double or triple, vascular occlusive lesions must be exceedingly widespread. But, occlusion assumes many forms: medial hypertrophy, fibrosis, arteritis, obliteration and amputation of arterial branches. Since the clinical picture of primary pulmonary hypertension may be a consequence of a variety of initiating mechanisms, it would be surprising if the pathologic changes proved to be consistent. And they are not (8, 11).

The "plexiform lesion," which some advocate as diagnostic of primary pulmonary hypertension, is also found in some instances of secondary pulmonary hypertension associated with congenital heart disease (5, 8, 11), in schistosomiasis (8), and in the pulmonary hypertension that coexists with hepatic cirrhosis (8, 44). The common denominator for this lesion is more apt to be a necrotizing arteritis—such as that produced by an episode of severe pulmonary hypertension—than a single etiologic agent (5, 7).

Few of the other vascular changes are as striking as the plexiform lesion. In some instances of primary pulmonary hypertension, intimal fibrosis is an outstanding feature; in a few medial hypertrophy predominates. Not infrequently, small emboli are found in peripheral branches; these seem to be secondary events.

The variability in the pulmonary vascular lesions from patient to patient cannot simply be dismissed as representing different stages in the evolution of a single process following some type of widespread vascular injury. Several reasons for the disparities other than differences in rates of progression can be suggested: (*a*) the duration and level of the pulmonary hypertension, (*b*) the age at which pulmonary hypertension begins and the state of the pulmonary circulation at the outset, (*c*) the sex of the patient, (*d*) the variable delay between the start of the vascular abnormalities and the first clinical manifestations, (*e*) the operation of different pathogenetic mechanisms, and (*f*) the coexistence of nonspecific vascular lesions that result from pulmonary hypertension from any cause with the vascular lesions triggered by the initiating agent.

HEMODYNAMICS

The hemodynamic findings in primary pulmonary hypertension are indistinguishable from pulmonary hypertension of other cause (9). Except for some of the aminorex patients (9), most patients are not subjected to catheterization until mean pulmonary artery pressures are in excess of 50 mm Hg; some have first been detected clinically when pulmonary arterial pressures exceeded 100 mm Hg. Right ventricular failure also has the usual features, i.e. an increase in right atrial and right ventricular diastolic pressure, often accompanied by a decrease in cardiac output. During exercise, the increase in pulmonary arterial pressure is inordinant for the increase in cardiac output, reflecting the consequences of the restricted pulmonary vascular bed.

Attention has also been called to spontaneous swings in pulmonary arterial pressure first noted in a young child with primary pulmonary hypertension (46). This observation supports the notion that in some individuals heightened vasomotor tone may contribute to the pulmonary hypertension; it also encourages the search for effective pulmonary vasodilators (4).

COURSE OF THE PULMONARY HYPERTENSION

Grim auguries for untreated primary pulmonary hypertension are traditional (47, 48). These appear warranted for the sporadic disorder that occurs in young or middle-aged women, especially when pulmonary hypertension is severe. But, the predictions are less reliable when applied to children or to males. The aminorex epidemic also has taught that some forms of the disorder either stabilize at a tolerable level or are reversible (9, 13). Indeed, as a general rule, the lower the pulmonary arterial pressure, the better is the prognosis (49). Also, it seems reasonable that predominant medial hypertrophy affords better prospect for reversal than do widely disseminated fibrous obliterative lesions.

VASODILATOR THERAPY

Both clinicians and pathologists strongly suspect that vasoconstriction contributes importantly to the pulmonary hypertension (3, 8, 50). But, in view of the heterogeneity of the pulmonary vascular lesions in disorders that are currently designated as primary pulmonary hypertension, it would be surprising if vasodilator therapy would be as uniformly successful in pulmonary, as in systemic, hypertension where medial hypertrophy is the predominant or the exclusive lesion. Unfortunately, experience has borne out this prediction: during the last 25 years, no therapeutic measures have proved consistently successful in relieving symptoms or affecting either the course or outcome of the disease (3, 4, 51). Chlorothiazide, reserpine and corticosteroids have been consistently unsuccessful (52); still under trial are a variety of ganglionic-blocking agents and vasodilators that appear to act directly on smooth muscle.

Before considering individual agents, two aspects that affect the interpretation of their effects warrant special mention. (*a*) The aim of therapy is to decrease the work of the right ventricle, primarily the afterload; but, the load may decrease not because of pulmonary vasodilation but because of other hemodynamic improvements, e.g. improved performance of the left ventricle. (*b*) The use of pulmonary "arteriolar" resistance as a measure of vasodilation is fraught with uncertainty unless either pulmonary vascular pressures or flow remains constant; if both are changing, the shape of pressure-flow relationships in the pulmonary circulation must be taken into account (46). In addition to these two problems in interpretation, there is also a practical problem in implementation: the results of the *acute* administration of a drug do not automatically predict the effects of *chronic* administration of the same agent, particularly if the route of administration is changed.

Oxygen

Attempts at pulmonary vasodilator therapy traditionally begin with the administration of 100% oxygen (51). The rationale is based on the capability of improved oxygenation to relieve hypoxic vasoconstriction, both acutely and chronically. However, unless appreciable arterial hypoxemia is present, the results are unimpressive.

Acetylcholine

Acetylcholine is a powerful pulmonary vasodilator when preexisting pulmonary vasomotor tone is high (3). Although repeated infusions over days in pulmonary hypertensive animals and children have succeeded in lowering pulmonary arterial pressure, repeated intravenous injections either at close intervals, or continuously, are impractical as a therapeutic measure for a chronic disorder. Therefore, acetylcholine probably has its greatest utility as a test for heightened pulmonary vasomotor tone.

Isoproterenol

Isoproterenol is a sympathomimetic amine that acts almost exclusively on β-receptors. In usual intravenous doses (up to 1 μg/kg), it increases cardiac output by inotropic and chronotropic effects it also increases venous return. Sublingual or oral administration is unreliable.

For acute studies, isoproterenol has been administered intravenously. Its effectiveness as an acute pulmonary vasodilator has been inconsistent (51, 53), and there has been no way, based on the antecedent clinical course, to predict its acute, or long-term, effectiveness.

The relationship between the variable therapeutic response to acute administration of isoproterenol and pulmonary vascular histology has not been examined. Unfortunately, pulmonary vasodilation in response to acute administration of the agent provides no assurance that the vasodilator effect will continue when either the oral or sublingual route is substituted for the intravenous route.

Tolazoline and Phentolamine

Tolazoline and phentolamine are 2-substituted imidazolines with a transient α-adrenergic-blocking effect. They also have important sympathomimetic, parasympathomimetic, and histamine-like effects. Tolazoline is now primarily of historical interest. Like isoproterenol, its effectiveness has been inconsistent (54). Phentolamine is a more potent adrenergic-blocking agent than tolazoline and its other effects are less prominent. Pulmonary arterial pressure and vascular resistance are often reduced transiently by tolazoline administered intravenously. In one recent case report of a 29-year-old man with pulmonary hypertension, phentolamine (250 mg per day in divided doses while awake) lowered pulmonary artery pressure and relieved symptoms during exercise for 20 months (54). However, as is the problem with most trials of vasodilator agents, it is difficult to generalize from a single experience, particularly when the patient is a male in whom the course of the illness is unpredictable and in whom the nature of the histological lesions in the pulmonary vascular tree is unknown.

CONCLUSION

The designation "primary pulmonary hypertension" refers to a variety of disorders. In attempting to identify the etiology, an important clue is the clinical setting in which the pulmonary hypertension arises. Because of the diversity in etiology and pathology, the prospects for achieving relief by a standard program of vasodilator therapy seem slim. This pessimism is borne out by the fact that to date exceedingly few instances of improvement have occurred in response to the use of pulmonary vasodilators. A much better prognosis is afforded if an inciting agent is identified and withdrawn, as in the case of aminorex pulmonary hypertension.

Literature Cited

1. Hatano, S., Strasser, T., eds. 1975. *Primary Pulmonary Hypertension. Report on a WHO Meeting.* Geneva: World Health Organization. 45 pp.
2. Fishman, A. P. 1979. Primary pulmonary hypertension. In *Pulmonary Diseases and Disorders,* ed. A. P. Fishman, 1:847–52. New York: McGraw-Hill. 1835 pp.
3. Harris, P., Heath, D. 1977. *Human Pulmonary Circulation. Its Form and Function in Health and Disease.* Edinburgh: Churchill Livingstone. 712 pp.
4. Voelkel, N. F., Reeves, R. J. 1979. Primary pulmonary hypertension. In *Pulmonary Vascular Disease,* ed. K. M. Moser, pp. 573–628. New York: Dekker
5. Fishman, A. P. 1975. Plexiform lesions. *Pathol. Microbiol.* 43:242–45
6. Tandon, H. D., Kasturi, J. 1975. Pulmonary vascular changes associated with isolated mitral stenosis in India. *Br. Heart J.* 37:26–36
7. Rüttner, J. R., Bartschi, P., Niederman, R., Schneider, J. 1979. Plexogenic pulmonary arteropathy and liver cirrhosis. *Thorax.* In press
8. Wagenvoort, C. A., Wagenvoort, N. 1977. *Pathology of Pulmonary Hypertension.* New York: Wiley. 345 pp.
9. Gurtner, H. P. 1979. Pulmonary hypertension, "plexogenic pulmonary arteropathy" and the appetite depressant drug aminorex: post or propter? *Bull. Eur. Physiopathol. Respir.* In press
10. Loogen, F., Both, A. 1976. Primary pulmonary hypertension. *Z. Kardiol.* 65:1–14
11. Widgren, S. 1977. Pulmonary hypertension related to aminorex intake. *Curr. Top. Pathol.* 64:1–64
12. Wagenvoort, C. A., Wagenvoort, N. 1970. Primary pulmonary hypertension. A pathologic study of the lung vessels in 156 clinically diagnosed cases. *Circulation* 42:1163–84
13. Both, A., Loogen, F., Maurer, W. 1971. Follow-up studies on patients with primary vascular pulmonary hypertension taking anorectics. *Verh. Dtsch. Ges. Inn. Med.* 77:445–48
14. Charters, A. D., Baker, W. C. 1970. Primary pulmonary hypertension of unusually long duration. *Br. Heart J.* 32:130–33
15. Bourdillon, P. D., Oakley, C. M. 1976. Regression of primary pulmonary hypertension. *Br. Heart J.* 38:264–70
16. Epler, G. R., McLoud, T. C., Gaensler, E. A., Mikus, J. P., Carrington, C. B. 1978. Normal chest roentgenograms in chronic diffuse infiltrative lung disease. *N. Engl. J. Med.* 298:934–39
17. Young, R. H., Mark, G. J. 1978. Pul-

monary vascular changes in scleroderma. *Am. J. Med.* 64:998–1004

18. Atkins, J. M., Mitchell, H. C., Pettinger, W. A. 1977. Increased pulmonary vascular resistance with systemic hypertension. *Am. J. Cardiol.* 39:802–7
19. Walloopillai, N. J. 1975. Primary pulmonary hypertension. An unexplained epidemic in Sri Lanka. *Pathol. Microbiol.* 43:248–50
20. Clinicopathologic Conference. 1976. Hypereosinophilic syndrome with pulmonary hypertension. *Am. J. Med.* 60:239–47
21. Kanemoto, N., Gonda, N., Katsu, M., Fukuda, J. 1975. Two cases of pulmonary hypertension with Raynaud's phenomenon. Primary pulmonary hypertension and systemic lupus erythematosus. *Jpn. Heart J.* 16:354–60
22. Williams, M. H., Adler, J. J., Colp, C. 1969. Pulmonary function studies as an aid in the differential diagnosis of pulmonary hypertension. *Am. J. Med.* 47:378–83
23. Kingdon, H. S., Cohen, L. S., Roberts, W. C., Braunwald, E. 1966. Familial occurrence of primary pulmonary hypertension. *Arch. Intern. Med.* 118:422–26
24. Rogge, J. D., Mishkin, M. E., Genovese, P. D. 1966. The familial occurrence of primary pulmonary hypertension. *Ann. Intern. Med.* 65:672–84
25. Czarnecki, S. W., Rosenbaum, H. M. 1968. The occurrence of primary pulmonary hypertension in twins with a review of etiological considerations. *Am. Heart J.* 75:240–46
26. Melmon, K. L., Braunwald, E. 1963. Familial pulmonary hypertension. *N. Engl. J. Med.* 269:770–75
27. Thompson, P., McRae, C. 1970. Familial pulmonary hypertension. Evidence of autosomal dominant inheritance. *Br. Heart J.* 32:758–60
28. Levin, D. L., Hyman, A. I., Heymann, M. A., Rudolph, A. M. 1978. Fetal hypertension and the development of increased pulmonary vascular smooth muscle: a possible mechanism for persistent pulmonary hypertension of the newborn infant. *J. Pediatr.* 92:265–69
29. Riggs, T., Hirschfeld, S., Fanaroff, A., Liebman, J., Fletcher, B., Meyer, R. 1977. Persistence of fetal circulation syndrome: an echocardiographic study. *J. Pediatr.* 91:626–31
30. Arias-Stella, J., Saldana, M. 1962. Muscular pulmonary arteries in people native to high altitude. *Med. Thorac.* 19:484–93
31. Haworth, S. G., Reid, L. 1976. Persistent fetal circulation: newly recognized structural features. *J. Pediatr.* 88:614–20
32. Peckham, G. J., Fox, W. W. 1978. Physiologic factors affecting pulmonary artery pressure in infants with persistent pulmonary hypertension. *J. Pediatr.* 93:1005–10
33. Rivier, J. L., Jaeger, M., Desbaillets, P., Reymond, C. 1971. Primary and anorexigenic pulmonary arterial hypertension. *Arch. Mal. Coeur Vaiss.* 64:607 (Abstr.)
34. Fishman, A. P. 1974. Dietary pulmonary hypertension. *Circ. Res.* 35:657–60
35. Byrne-Quinn, E., Grover, R. F. 1972. Aminorex (Menocil) and amphetamine: acute and chronic effects on pulmonary and systemic haemodynamics in the calf. *Thorax* 27:127–31
36. Edwards, W. D., Edwards, J. E. 1977. Clinical primary pulmonary hypertension. Three pathologic types. *Circulation* 56:884–88
37. Manchester, D., Margolis, H. S., Sheldon, R. E. 1976. Possible association between maternal indomethacin therapy and primary pulmonary hypertension of the newborn. *Am. J. Obstet. Gynecol.* 126:467–69
38. Kay, J. M., Heath, D. 1969. *Crotalaria Spectabilis: Pulmonary Hypertension Plant.* Springfield, Ill: Thomas. 146 pp.
39. Atkinson, B. F., Katz, A. S., Knight, D., Pietra, G. G., Fishman, A. P. 1977. Monocrotaline-induced pulmonary hypertension in beagle puppies. *Lab. Invest.* 36:354 (Abstr.)
40. Raczniak, T. J., Shumaker, R. C., Allen, J. R., Will, J. A., Lalich, J. J. 1979. Pathophysiology of dehydromonocrotaline-induced pulmonary fibrosis in the beagle. *Respiration* 37:252–60
41. Fahlén, M., Bergman, H., Helder, G., Rydén, L., Wallentin, I., Zettergren, L. 1973. Phenformin and pulmonary hypertension. *Br. Heart J.* 35:824–28
42. Heath, D., Smith, P. 1977. Pulmonary vascular disease. *Med. Clin. North Am.* 61:1279–1307
43. Senior, R. M., Britton, R. C., Turino, G. M., Wood, J. A., Langer, G. A., Fishman, A. P. 1968. Pulmonary hypertension associated with cirrhosis of the liver and with portacaval shunts. *Circulation* 37:88–96
44. Naeye, R. L. 1960. "Primary" pulmonary hypertension with coexisting portal hypertension. A retrospective study of six cases. *Circulation* 22:376–84

45. Daoud, F. S., Reeves, J. T., Schaefer, J. W. 1972. Failure of hypoxic pulmonary vasoconstriction in patients with liver cirrhosis. *J. Clin. Invest.* 51:1076–80
46. Fishman, A. P. 1963. Dynamics of the pulmonary circulation. In *Handbook of Physiology. Circulation,* ed. W. F. Hamilton, P. Dow, 2:1667–743. Washington, DC: American Physiological Society. 1786 pp.
47. Wade, G., Ball, J. 1957. Unexplained pulmonary hypertension. *Q. J. Med.* 26:83–119
48. Wood, P. 1958. Pulmonary hypertension with special reference to the vasoconstrictive factor. *Br. Heart J.* 20: 557–70
49. Voss, H., Harms, H. 1970. Epidemiology and clinical aspects of primary vascular pulmonary hypertension. A report on 42 cases. *Z. Kreislaufforsch.* 59:887–91
50. Walcott, G., Burchell, H. B., Brown, A. L. Jr. 1970. Primary pulmonary hypertension. *Am. J. Med.* 49:70–79
51. Daoud, F. S., Reeves, J. T., Kelly, D. B. 1978. Isoproterenol as a potential pulmonary vasodilator in primary pulmonary hypertension. *Am. J. Cardiol.* 42:817–22
52. Sleeper, J. C., Orgain, E. S., McIntosh, H. D. 1962. Primary pulmonary hypertension: review of clinical features and pathologic physiology with a report of pulmonary hemodynamics derived from repeated catheterization. *Circulation* 26:1358–69
53. Shettigar, U. R., Hultgren, H. N., Specter, M., Martin, R., Davies, D. H. 1976. Primary pulmonary hypertension. Favorable effect of isoproterenol. *N. Engl. J. Med.* 295:1414–15
54. Rushkin, J. N., Hutter, A. M. Jr. 1979. Primary pulmonary hypertension treated with oral phentolamine. *Ann. Intern. Med.* 772–74

Ann. Rev. Med. 1980. 31:433–39
Copyright © 1980 by Annual Reviews Inc. All rights reserved

FOLATE-BINDING PROTEINS ❖7371

Neville Colman, M.D., Ph.D. and Victor Herbert, M.D., J.D.

Hematology and Nutrition Laboratory, Bronx Veterans Administration Medical Center, Bronx, New York 10468; and Department of Medicine, State University of New York—Downstate Medical Center, Brooklyn, New York 11203

Current studies of proteins that specifically bind folic acid and its analogues with very high affinity can be traced to the reports (1, 2) that milk contained endogenous *Lactobacillus casei*-active folate bound to a heat labile binder which could also bind exogenous folate but little or no methotrexate. Studies using ^{3}H-folic acid confirmed that it remained bound in the presence of activated charcoal, and demonstrated that the binding was saturable, virtually abolished at pH 3.5, and not significantly affected by competition with either reduced folates substituted at *N*-5 (5-methyltetrahydrofolate, folinic acid) or 4-amino analogues (aminopterin, methotrexate). The studies also showed that bound folate eluted at low ionic strength from an ion-exchange column (3). The milk folate binder was a discrete minor whey protein that eluted from gels with beta lactoglobulin with an apparent molecular weight of 38,000 and showed concentration-dependent reversible aggregation (4). The protein was purified using affinity chromatography (5) and characterized as a glycoprotein by several groups. It may impair uptake of folate into enteric microorganisms (6), thereby allowing selective delivery of milk folate to the intestinal mucosa. The main practical effect of these studies has been in developing folate radioassays, in which milk remains the most widely used source of folate binder (7–9).

GRANULOCYTE FOLATE BINDER

Granulocytes from a minority of subjects contain protein that binds easily detectable amounts (>20 pg) of ^{3}H-folic acid. This was discovered by chance when it inactivated, by binding, the tracer substrate being used to study dihydrofolate reductase in chronic granulocytic leukemia cells (10). In preliminary studies, the characteristics of this binder were essentially identical to those of milk folate binder. Later, an immunochemically identical factor with similar binding characteristics was found in leukocyte

0066-4219/80/0401-0433$01.00

lysates of women who were pregnant or taking oral contraceptives (11). Thus, this factor was known not to be specifically associated with the leukemia; but reports of failure to find it in normal controls preserved the concept that it was an aberrant manifestation (11).

Chronic granulocytic leukemia cells yielded folate-binding activity that resolved into two distinct fractions on both ion exchange and gel exclusion chromatography. Whereas the larger molecule bound folate irreversibly, binding of reduced but not oxidized folate by the low molecular weight protein was partly reversible, which suggested a possible transport function (12). After purification by affinity chromatography, both binders migrated together as a single band in polyacrylamide disc gel electrophoresis and sedimentation equilibrium centrifugation, and showed immunologic identity (13). A difference in isoelectric points was apparently due to differing sialic acid content. In one study, two of four subjects with binder in chronic granulocytic leukemia cells demonstrated avid binding of methotrexate (14). If this observation is confirmed and if similar binders are found in subjects liable to receive methotrexate therapy, it may portend significantly altered pharmacokinetics of this agent. The binder is synthesized in the leukemic cells by mechanisms that are temperature dependent, and sensitive to ribosomal protein inhibition and sulfhydryl inhibitors (15).

These binders were detected in up to 30% of normal controls' granulocytes during kinetic studies of release in different population groups (16). Granulocyte release of these binders was similar to that of vitamin B_{12} binders (17) in that those granulocyte populations that contained folate binder released this in the presence of 50-mM lithium chloride or a mixture of calcium and the ionophore A23187.

The latter agent can stimulate neutrophils to release more than 90% of specific granule markers, such as transcobalamin, and less than 10% of azurophilic granule markers such as β-glucuronidase. Studies using this tool indicated that granulocyte folate binders are almost exclusively localized within the specific granules of these cells (18), despite the prior report of subcellular localization in the heavy nuclear and mitochondrial fractions (19). The release observed in the presence of ionophore A23187 plus calcium and that in the presence of lithium had similar functional and kinetic characteristics, which generally approximated those that might be observed during phagocytosis.

Although a functional role for folate binder in granulocytes has not been demonstrated, their release under conditions mimicking phagocytosis suggests that they may play a role in primary cellular responses to injury. Since the binding reaction withholds folate from bacterial colonies (6), such a role could include control of bacterial proliferation by scavenging folate required for *de novo* synthesis of thymidine-DNA.

SERUM FOLATE BINDER

It is important to differentiate specific, high affinity, serum folate binding from the weak, nonspecific association that albumin and other serum proteins form with almost half the folate in serum. High affinity binders were first reported in sera of some patients with chronic granulocytic leukemia (20) and subsequently in alcoholics with folate deficiency (21), uremics (22), and subjects with liver disease (23). Detectable amounts of unsaturated folate binder are present in approximately 15% of normal subjects (24).

The family tree of a group of normal subjects with elevated folate binder suggested an autosomal dominant mode of inheritance of this characteristic (25). Treatment of serum with charcoal at acid pH strips endogenous-bound folate from its binding sites, which can then be labeled at neutral pH with tracer folate (24). Through the use of this technique, high affinity folate binders have been demonstrated in all human and animal sera tested. Removal of endogenous folate at acid pH by dialysis or coated-charcoal exposure yielded unsaturated binding sites suitable for Scatchard plot analysis, which suggested a single class of binders with an association constant for folic acid 2.8×10^{10} liters per mole. Thus, only 16 ng of folic acid per liter are required to half-saturate these binding sites, although almost all serum folate is 5-methyltetrahydrofolate, which is bound with much lower affinity.

The serum folate binder elutes as a single peak in gel exclusion chromatography corresponding to an apparent molecular weight of 42,000 (24). Although serum folate binder previously appeared homogeneous in DEAE cellulose chromatography, we were able to elute two discrete peaks from DEAE-Sephadex® (26), the first with phosphate-buffered saline at 0.08 M and consisting almost entirely of unsaturated binder, and the second appearing at 0.12 M with equivalent amounts of saturated and unsaturated binder (27).

This chromatographic evidence for heterogeneity was supported by the observation that serum folate binder of probable granulocytic origin bound folic acid optimally at a lower pH and with less inhibition by methyltetrahydrofolate than that of probable hepatic origin (27).

Many properties of the serum binders are similar to those of granulocytes, including rapid association and slow dissociation of bound folic acid, preference for oxidized rather than reduced folate, elution as a β-globulin, recovery in the transferrin band after polycrylamide gel electrophoresis, and an apparent molecular weight of about 40,000 (28). The acid dissociation technique (23) was used to remove bound folate from normal neonatal cord serum, uncovering a high affinity binder similar in molecular weight to that observed in adults, but showing a less marked preference for folic acid over methyltetrahydrofolate, which exhibited equal competition when present in a molar ratio of 1:4 (29).

Alterations occurring in different clinical conditions were assessed by measuring unsaturated binding capacity (UFBC), total binder (TFBC), and percentage saturation in normal subjects of both sexes, azotemic males, males with histologically proven cirrhosis, and pregnant women. The results are shown in Table 1. All normal subjects had greater than 33% saturation of the binder, with an average saturation of 67%. The observed range of TFBC was 100–325 pg/ml, with 3 of 20 showing UFBC greater than 125 pg/ml. Total binder levels were above this range in 40% of azotemic males, in whom percentage saturation was near the normal range. Thus a normal binder with normal affinity may be present in increasing amounts, perhaps because of an interference with normal urinary excretion. In contrast, in none of the cirrhotic subjects were total binder levels elevated, but unsaturated binder was significantly elevated in 60%, which suggests that the protein was present in normal quantities but had altered affinity for methyltetrahydrofolate compared with folic acid. Finally, pregnant women demonstrated two abnormal patterns: 27% had grossly elevated total binder levels with saturation decreased below 33%, whereas a proportion of those with normal total binder levels also had decreased saturation (24).

Fernandes-Costa & Metz (30) reported that binding sites for methyltetrahydrofolate in sera of pregnant women were fully saturated in the physiological state and that binding of added ^{3}H-folic acid was mainly to sites not available to methyltetrahydrofolate. This suggested a mechanism for the previously observed differences in competition of methyltetrahydrofolate for folic acid binding in different sera (27, 29). They also suggested that binding of methyltetrahydrofolate and/or folic acid altered the configuration of the binding protein sufficiently to change the elution pattern in gel exclusion chromatography (30).

Table 1 Unsaturated (UFBC) and total (TFBC) folate-binding capacity and percentage saturation (Mean ± SE)

Patients	Number of patients	UFBC (pg/ml)	TFBC (pg/ml)	Saturation (%)
Normal	20	61 ± 11	174 ± 16	67 ± 5
Uremic	20	119 ± 18	344 ± 44	64 ± 5
Cirrhotic	10	131 ± 32	226 ± 30	45 ± 9
First trimester of pregnancy	25	131 ± 40	248 ± 46	61 ± 5
Second trimester of pregnancy	49	207 ± 33	299 ± 36	43 ± 3
Third trimester of pregnancy	20	155 ± 52	257 ± 49	53 ± 7

The role of high affinity serum folate binders in cellular folate transport remains obscure. It can impair the uptake of folic acid and methyltetrahydrofolate into human bone marrow cells in vitro (31) and into HeLa cell cultures in vitro (32). A folate derivative linked to amino acid labeled with gamma-emitting selenium and coupled to pregnancy serum folate binder was delivered predominantly to the liver in both pregnant and nonpregnant animals, in contrast to significant uptake into placenta and fetal tissues when weakly bound to nonspecific serum proteins (33).

OTHER FLUIDS AND TISSUES

Cerebrospinal fluid contains folate binder identical to that in serum except that it is primarily unsaturated (34). Similar binders have been noted in urine and bile but have not been well characterized.

The first folate binder identified in other body tissues was that partially purified by ethanol fractionation of a citric acid extract of hog kidney acetone powder. It resembled granulocyte and milk binders in its relative affinity for the different folates, poor binding of methotrexate and folinic acid, dissociation at pH 3.5, and reversibility of binding below pH 6 (35). Although binding by the acetone powder was essentially irreversible and independent of folate concentration above pH 6, subsequent study of fresh rat kidney, in which brush border membrane was the main source of binder, showed binding kinetics virtually identical to those of serum, with an association constant of 2.4×10^{10} liters per mole (36). It was hypothesized that brush border binder might contribute to the reabsorption of folate from glomerular filtrate.

Different techniques have been used to study folate binder in liver by following the localization and behavior of radioactivity after intraperitoneal injection of ^{3}H-folic acid (37). Four binders of different molecular weight in the cytosol and mitochondria were identified, binding primarily polyglutamate folate. These fractions bound significant amounts of tritiated polyglutamates but small amounts of monoglutamates in vitro. The data suggested the existence of a membrane carrier for radioactive folic acid on the plasma membrane fraction of the liver cells, with probable storage of natural folates in one of the cytosol-binding proteins (37).

Folate binding by a probable membrane carrier was identified in solubilized materials from the particulate fraction of choroid plexus homogenate, with an affinity constant similar to that in serum (38). Brush border membrane proteins of the epithelial cells of the small intestine bound folate differently from all the tissues previously noted, having a much lower affinity and the capacity for stereospecific inhibition by *p*-aminobenzoylglutamate, in contrast to all binders previously noted (39). A membrane carrier function for this protein has not been demonstrated.

CONCLUSION

It appears certain that only a small part of the spectrum of folate binder action has been revealed as yet, but with the data at hand several possible modes of action can be anticipated. First, milk folate-binding protein may withhold the vitamin from enteric bacteria and may even directly enhance mucosal transport of the vitamin across the wall of the small intestine (40). Second, the binder in granulocytes may be released during phagocytosis and play a role in the control of infection by preventing invading microorganisms from synthesizing DNA. Third, serum folate binder may facilitate selective transport of certain folates into dividing cells and other folates into liver. Cerebrospinal fluid binder may be important in transport of folate in the central nervous system, whereas the binder located in the choroid plexus is the archetype of intracellular folate binders, which may function as membrane carriers and/or as mechanisms for the storage of folates of appropriate chemical configuration.

Literature Cited

1. Ghitis, J. 1966. The labile folate of milk. *Am. J. Clin. Nutr.* 18:452
2. Ghitis, J. 1967. The folate binding in milk. *Am. J. Clin. Nutr.* 20:1
3. Metz, J., Zalusky, R., Herbert, V. 1968. Folic acid binding by serum and milk. *Am. J. Clin. Nutr.* 22:289–97
4. Ford, J. E., Salter, D. N., Scott, K. J. 1969. The folate-binding protein in milk. *J. Diary Res.* 36:435
5. Salter, D. N., Ford, J. E., Scott, K. J., Anderson, P. 1972. Isolation of the folate binding protein from cows milk by the use of affinity chromatography. *FEBS Lett.* 20:302
6. Ford, J. E., Scott, K. J., Sansom, B. F., Taylor, P. J. 1975. Some observations on the possible nutritional significance of vitamin B_{12}- and folate-binding proteins in milk. Absorption of (^{58}Co)-cyanocobalamin by suckling piglets. *Br. J. Nutr.* 34:469
7. Waxman, S., Schreiber, C., Herbert, V. 1971. Radioisotopic assay for measurement of serum folate levels. *Blood* 38:219–28
8. Rothenberg, S. P., daCosta, M., Rosenberg, Z. 1972. A radioassay for serum folate: Use of a two-phase sequential-incubation, ligand-binding system. *N. Engl. J. Med.* 286:1335–39
9. Longo, D. L., Herbert, V. 1976. Radioassay for serum and red cell folate. *J. Lab. Clin. Med.* 87:138–51
10. Rothenberg, S. P. 1970. A macromolecular factor in some leukemic cells which binds folic acid. *Proc. Soc. Exp. Biol. Med.* 133:428
11. daCosta, M., Rothenberg, S. P. 1974. Appearance of a folate binder in leukocytes and serum of women who are pregnant or taking oral contraceptives. *J. Lab. Clin. Med.* 83:217–14
12. Fischer, C. D., daCosta, M., Rothenberg, S. P. 1975. The heterogeneity and properties of folate binding proteins from chronic myelogenous leukemia cells. *Blood* 46:855–67
13. Fischer, C. D., daCosta, M., Rothenberg, S. P. 1978. Properties of purified folate-binding proteins from chronic myelogenous leukemia cells. *Biochim. Biophys. Acta* 543:328–39
14. Corrocher, R., Pachor, M. L., Bambara, L. M., Ambrosetti, A., DeSandre, G. 1978. Folic acid binding protein in chronic granulocytic leukemia: The effect of methotrexate. *Acta Haematol.* 60:341–49
15. daCosta, M., Rothenberg, S. P. 1976. Studies with the folate binding protein in chronic granulocytic leukemia cells,

I. Synthesis and release of binder by cells in short-term culture. *Br. J. Haematol.* 34:581–87

16. Colman, N., Herbert, V. 1974. Release of folate binding protein from granulocytes. *Program, 17th Ann. Meet. Am. Soc. Hematol.,* p. 155
17. Scott, J. M., Bloomfield, F. J., Stebbins, R., Herbert, V. 1974. Studies on derivation of Transcobalamin III from granulocytes. Enhancement by lithium and elimination by fluoride of *in vitro* increments in vitamin B_{12} binding capacity. *J. Clin. Invest.* 53:228–29
18. Colman, N., Herbert, V. 1979. Studies using the calcium ionophore A23187 suggest localization of the human granulocyte folate binder in specific (secondary) granules. *Clin. Res.* 27: 291A
19. daCosta, M., Rothenberg, S. P. 1974. Studies of the folate binding factor in cultures and subcellular fractions of chronic myelogenous leukemia (CML) in cells. *Clin. Res.* 22:486A
20. Rothenberg, S. P., daCosta, M. 1971. Further observations on the folate binding factor in some leukemic cells. *J. Clin. Invest.* 50:719–26
21. Waxman, S., Schreiber, C. 1973. Measurement of serum folate levels and serum folic acid-binding protein by ^{3}H-PGA radioassay. *Blood* 42:281–90
22. Hines, J. D., Kamen, B., Caston, D. 1973. Abnormal folate binding protein(s) in azotemic patients. *Program, 16th Ann. Meet., Am. Soc. Hematol.,* p. 57
23. Colman, N., Herbert, V. 1974. Measurement of total folate binding capacity (TFBC) of serum after removal of endogenous folate at acid pH. *Clin. Res.* 22:700A
24. Colman, N., Herbert, V. 1976. Total folate binding capacity of normal human plasma and variations in uremia, cirrhosis, and pregnancy. *Blood* 48:911–12
25. Muckerheide, M., Racih, P. C., Carlson, I. H. 1977. Increased serum folate binding capacity. *Acta Haematol.* 58:45–49
26. Colman, N., Herbert, V. 1978. DEAE Sephadex chromatography of human serum folate binding proteins. *Fed. Proc.* 37:493
27. Colman, N., Herbert, V. 1979. Kinetic and chromatographic evidence for heterogeneity of the high affinity folate binding proteins in serum. In *Chemistry and Biology of Pteridines,* ed. R. Kisliuk, G. Brown, pp. 525–30, New York: Elsevier/North Holland. 713 pp.
28. Waxman, S., Schreiber, C. 1973. Characteristics of folic acid-binding protein in folate deficient serum. *Blood* 42:291–301
29. Kamen, B. A., Caston, J. D. 1975. Purification of folate binding factor in normal umbilical cord serum. *Proc. Natl. Acad. Sci.* 72:4261–64
30. Fernandes-Costa, F., Metz, J. 1979. Binding of methylfolate and pteroylglutamic acid by the specific serum folate binder. *J. Lab. Clin. Med.* 93: 181–88
31. Tisman, G., Herbert, V. 1971. Inhibition by human serum, dilantin (diphenylhydantoin), and methotrexate (MTX), and enhancement by 2-deoxyglucose, of ^{3}H-pteroylglutamic acid (^{3}H-PGA) and ^{3}H-5-methyltetrahydrofolate (M-THF) uptake by human bone marrow cells *in vitro. Clin. Res.* 19:433
32. Waxman, S., Schreiber, C. 1974. The role of folic acid binding proteins in the cellular uptake of folates. *Proc. Soc. Exp. Biol. Med.* 147:760–64
33. Rubinoff, M., Abramson, R., Schreiber, C., Waxman, S. 1977. Differences in folate transport and distribution caused by folate binding protein (FABP). *Clin. Res.* 25:317A
34. Colman, N., Herbert, V. 1977. Folate binders in cerebrospinal fluid. *Clin. Res.* 25:366A
35. Kamen, B. A., Caston, J. D. 1975. Identification of a folate binder in hog kidney. *J. Biol. Chem.* 250:2203–5
36. Selhub, J., Rosenburg, I. H. 1978. Demonstration of high-affinity folate binding activity associated with the brush border membranes of rat kidney. *Proc. Natl. Acad. Sci.* 75:3090–93
37. Zamierowski, M. M., Wagner, C. 1977. Identification of folate binding proteins in rat liver. *J. Biol. Chem.* 252:933–38
38. Spector, R. 1977. Identification of folate binding macromolecular in rabbit choroid plexus. *J. Biol. Chem.* 252:3364–70
39. Leslie, G. I., Rowe, P. B. 1972. Folate binding by the brush border membrane proteins of small intestinal epithelial cells. *Biochemistry* 11:1696–1703
40. Herbert, V., Hettiarachchy, N., Colman, N. 1979. Evidence for existence of an "intrinsic factor" for folate in milk which can deliver folate to mammalian intestine receptors. *Blood* 54 (No. 5, Suppl. 1), 39A

Ann. Rev. Med. 1980. 31:441–51
Copyright © 1980 by Annual Reviews Inc. All rights reserved

ACTIVE IMMUNIZATIONS FOR ADULTS

❖7372

Gilbert M. Schiff, M.D.

The Christ Hospital Institute of Medical Research and the University
of Cincinnati College of Medicine, Cincinnati, Ohio 45219

Lack of effective antiviral drugs has led to dependence on vaccines for control of many viral diseases. These diseases are primarily "childhood diseases," so immunization activities have largely been directed toward children. Notable exceptions to this emphasis on children are influenza immunizations. Even with effective viral vaccines, significant efforts to immunize children, and generous financial support to purchase vaccines, it is becoming evident that a portion of the adult population remains susceptible to one or more of these common childhood diseases. These adults are "victims" of one or more of the following circumstances: (*a*) the widespread use of vaccines creates a herd immunity that reduces the usual exposure to these viruses; (*b*) not all persons are immunized in childhood; or (*c*) some persons receive ineffective or mishandled vaccines (vaccine failure). The bottom line is that an increasing proportion of adults could greatly benefit from "childhood" vaccinations. The situation is somewhat complicated by the relative lack of knowledge of these vaccines among nonpediatricians, as well as indications that these vaccines are less antigenic (1) and at times more reactive (2) in adults.

Adults are the primary target for influenza and rabies vaccines, although children frequently receive the latter. Influenza immunization creates confusion, frustration, and/or fright because of the many influenza strains that have been isolated, the phenomena of antigen drift and shift, the unpredictability of influenza outbreaks, the varying manufacturing processes, the creation of "priority" categories to receive the vaccines, frequent reactions, and the unexplained relation between "Swine" influenza vaccine and the Guillain-Barré syndrome. Decisions on administration of rabies vaccination may be delayed and the vaccine may cause serious side effects.

Several viral vaccines are required only by the traveler, and at times are unavailable at the family practitioner's or internist's office. Finally, in the

0066-4219/80/0401-0441$01.00

area of viral research, newer vaccines that will involve adults are under development.

There are several bacterial vaccines available for adults. Some are used only because of foreign travel requirements and have dubious effectiveness. Diphtheria toxoid is occasionally given to an adult, and wound management frequently involves tetanus toxoid administration in adults. New capsular polysaccharide bacterial vaccines are indicated for certain segments of the adult population.

There are two rickettsial vaccines: typhus and Rocky Mountain Spotted Fever vaccines. The former is used in travelers, and the latter recommended for a select adult population only.

Before consideration of the individual vaccines, certain statements should be made:

1. In general, live, attenuated viral vaccines are superior antigens to killed (inactivated) viral vaccines in original antibody response, booster response, and duration of antibody response.
2. Except for live triple oral polio vaccine, the risk of contact infection from live virus vaccines is almost nil. (This refers to usual, nonplacental, modes of spread.)
3. Except for rubella administration to a woman who is pregnant or gets pregnant within three months, vaccination is not contraindicated in pregnancy if the benefit from the vaccination outweighs the risk of contracting the natural disease. However, in general, live vaccines should be avoided whenever possible.
4. The accurate determination of immunity to the viral diseases is by serological test. "Significant" or "protective" serum antibody levels will vary from laboratory to laboratory.
5. No vaccine results in a 100% immunity rate. Where practical, and important enough, follow-up serology should be conducted post-vaccination.
6. Reasons for vaccine failure include bad lot, mishandling of lot, viral interference, concomitant administration of immune serum globulin, and, in the infant, presence of maternal antibody.
7. Live vaccines are generally contraindicated in immunosuppressed and immunodeficient individuals.

VIRAL VACCINES

Influenza Vaccines

Many virologists have spent lifetimes on trying to control, if not eradicate, influenza. Man is beset by three types of influenza: A, B, and C. Influenza A occurs in endemic, epidemic, and pandemic fashion, has a high morbidity, and, at times, a high mortality rate. It has been found worldwide, and tends to occur in the cold months of the year. While usually a self-limited illness confined to the upper respiratory tract and associated with significant systemic signs and symptoms, complications may occur. The most common are the pulmonary complications, which are pure influenza infiltration, secondary bacterial infection, or both. These complications usually have severe consequences. Influenza A occurs in both adults and children. Influenza vaccines have been made since the 1940s. But the occurrence of frequent antigenic drift and less frequent antigenic shift has compromised Influenza A vaccine-induced protection. To be most effective, Influenza A vaccines should contain the contemporary strains of Influenza A circulating in the population, should be given in adequate dose at least two weeks before exposure to the virus, and should be given as an annual booster. If these criteria are met, influenza vaccines in civilian populations are 60–70% effective (3, 4). Influenza B also occurs in endemic, epidemic, and pandemic fashion but to a lesser degree than Influenza A. Influenza B undergoes less frequent and less significant antigenic shift. Influenza B appears to affect children more than adults. The same rules for Influenza B vaccines apply as for the A strain vaccines. Influenza C is a very uncommon human pathogen, and no Influenza C vaccine exists.

All influenza vaccines licensed in the United States are egg-grown, inactivated vaccines. The vaccines are usually polyvalent (contain one or more A and one B strain), and are manufactured as "whole" or "split" product vaccines. The former essentially contains the whole, purified virus, while the latter contains the purified antigens responsible for inducing protective antibody. Recent improvement in manufacturing procedures has resulted in purified preparations with relatively increased amounts of desired antigens. This gives rise to fewer side effects and better antigenic activity. There is some debate about whether the "whole" or the "split" product vaccine is superior in terms of antigenicity and reactivity (5, 6). Dosage in adults varies with age. The general recommendation is a single inoculation for persons over 27 years of age with either whole or split product vaccine, and two inoculations 2–4 weeks apart of split product vaccine for persons 18–26 years old.

Reactions to influenza vaccines are common but usually mild to moderate. Most frequent complaints are soreness and redness at site of inoculation. At times severe reactions can occur. Persons with a known allergy to eggs or egg products should not receive influenza vaccine.

The apparent association of Guillain-Barré syndrome with Influenza A/76/NJ (Swine) vaccine remains an enigma. While opinion differs, it is generally thought that influenza vaccine should be given to the high risk groups (see below).

Much effort has been devoted to the development of live, attenuated influenza vaccines (7, 8). Several schemes have been devised to manufacture them, including temperature-sensitive mutant hybrids, cold-adapted hybrids, and serum inhibitor resistant strains. The hoped for advantages of live influenza vaccines are (*a*) rapid incorporation of new antigenic strains with rapid manufacture; (*b*) ease of administration (orally); (*c*) stimulation of local IgA antibody; (*d*) persistence of immunity; and (*e*) reduction of serious side effects. Several potential live, attenuated strains have been tested in humans. (9) Our laboratory in Cincinnati field-tested the serum inhibitor hybrid Alice (A) and Barbara (B) strains. The A strain was very effective (90%) (10) and the B strain was ineffective (40%), as was the combination (A-50%, B-30%) in high school children. Obviously the B strain interfered with the A.

The policy for administration of influenza vaccines in this country sets up priority categories. These categories consist of persons who by history are either more susceptible to influenza and/or have a higher incidence of severe disease and complications. Such groups are the elderly (over 55 years), persons with chronic pulmonary and cardiac diseases, persons immunosuppressed by disease or drug, persons in crowded conditions, medical and hospital personnel, and all essential personnel.

Several anti-influenzal drugs have recently been or are currently being developed. Amantadine has been licensed and its congener rimantadine has been used successfully in prophylaxis against many Influenza A strains (11, 12). However, daily administration of drug must be maintained throughout the duration of the epidemic. There is some evidence that these drugs may have a therapeutic effect (13). My personal recommendation for amantadine is to use it prophylactically during an Influenza A outbreak on an individual basis until two weeks after the proper vaccine can be administered.

Another promising anti-influenzal drug is Isoprinosine ®. Several recent, well-controlled human trials in the US have indicated favorable therapeutic effect by the drug against Influenza A strains (14–16). This drug is not licensed in the US.

Rubella Vaccines

Rubella vaccines (17, 18) were licensed for use in the US in 1968–1969. These were three live, attenuated virus vaccines that differed in the type of cell line on which the virus was grown. All gave a 95–98% antigenic response when administered to susceptible children and adults. It soon became evident that one vaccine (HPV-77, DK-12) produced unacceptable side effects (parethesias, arthralgia, arthritis); it was withdrawn from the market. A newer vaccine, RA 27/3, grown in human cells (WI-38), has recently been licensed in the US (19).

The strategy for use of rubella vaccine in this country has been to vaccinate all children 12 months of age until puberty. In this manner, an effective herd immunity might be produced that would prevent natural rubella outbreaks from flourishing, and hence the real targets of rubella prophylaxis, the susceptible pregnant women, might receive indirect protection. Because the vaccines were live viral products, a theoretical danger of vaccine-induced teratogenesis existed if given to pregnant women. After millions of doses of rubella vaccine have been given, the situation is as follows: (*a*) the incidence of congenital rubella dramatically decreased; (*b*) no vaccine-induced defective baby has been born, although abortuses of women vaccinated in early pregnancy yielded vaccine strain virus (20); (*c*) the herd immunity achieved has not stopped rubella outbreaks; (*d*) the outbreaks of rubella are occurring in persons not vaccinated, i.e. older teenagers, college students, young adults; (*e*) vaccine-induced immunity is slowly decreasing with time (10% are no longer protected at 12 years) (21); (*f*) teenage females appear to lose their immunity at a faster rate than younger females (1); (*g*) 20–35% of 18–25-year-old females remain susceptible (22); (*h*) history of previous clinical illness is worthless as a determinant of immunity; (*i*) the rubella vaccine of choice is the RA 27/3 vaccine (19).

What all this means is that 20–35% of today's child-bearing-age females are susceptible to rubella, and rubella is still occurring in this age group. Some attempt should be made to immunize these women. To this end, we do the following (23): (*a*) Recommend serological testing for every woman of child-bearing potential. (*b*) Vaccinate as many of those that are susceptible as possible on an individual basis. We educate the women to the danger of being pregnant at the time of vaccination or of becoming pregnant for three months afterward. We encourage some type of birth control during the three-month period, and selectively vaccinate women (i.e. not all are on contraceptives). (*c*) We retest the vaccinees 6–8 weeks after vaccination to ensure the presence of antibody. (In our experience with young adult women, vaccine failure is as high as 20%.) If there is vaccine failure, we revaccinate, and then retest.

Our studies show that 15–20% of young adult males are also susceptible. We encourage their vaccination because natural rubella in any adult can result in unpleasant and persistent joint and muscle symptoms. We also stress the fact that the unvaccinated, susceptible male may spread the virus to the pregnant woman.

Only a single inoculation of rubella vaccine is needed unless follow-up antibody assay indicates failure. The side effects of the RA 27/3 vaccine occur in about 5% of the population; they are usually mild, but can occasionally cause an arthralgia or arthritis. The latter may resemble rheumatoid arthritis but is temporary and nonresistant.

Measles Vaccine

Measles vaccines (24, 25) became available in the mid 1960s. Initially, there were inactivated and live, attenuated vaccines. A dose of immune serum globulin was recommended with the live vaccine. Soon, further attenuated measles vaccines were developed and are now the current vaccines of choice. Measles vaccination has become part of routine well-baby care.

In the past, measles rarely occurred in adults and vaccination was not considered for this age group. The situation has changed probably as a result of the measles vaccination program and its effects on "measles ecology." In the US, large epidemics of measles are occurring among the older teenagers, college students, and young adults (26). Mixed in with these epidemics are cases of atypical measles in persons who were given inactivated measles vaccines in the past (27). It appears that a significant portion of young adults currently lack measles immunity. They escaped the "immunization network" and natural exposure was reduced because of the community effects of effective vaccination programs. Adult measles is frequently a more serious disease than childhood measles (27). It is reasonable to immunize young adults who are considered susceptible by history alone with further attenuated measles vaccine (a single inoculation). Side reactions occur in 15% of the vaccinees and are usually mild. There is no danger in immunizing persons who already possess antibody.

Mumps Vaccine

Live, attenuated mumps vaccine (28) became available in the 1960s. Unlike measles and rubella vaccines, there has not been uniform acceptance of the routine use of mumps vaccine in the pediatric population. Despite a 95% seroconversion rate and minimal side-effects from the vaccine, some physicians were dubious about the persistence of vaccine-induced immunity, and preferred to permit their patients to be exposed to natural mumps with its resultant long-term immunity.

The problem of mumps in the adult concerns the more serious nature of the disease (i.e. orchitis, oophoritis) in post-pubertal persons. Physicians frequently encounter the father who has no history of clinical mumps or vaccination and has children with acute cases of mumps. The use and effectiveness of mumps hyperimmune globulin is not clear.

Mumps vaccine may be given to susceptible adults. Susceptibility may be determined by a serum neutralization test, but the test is relatively difficult to perform and is not readily available. The mumps skin tests are totally unreliable. It is reasonable and practical to rely on lack of history of clinical mumps for a decision on whether or not to administer mumps vaccine. Subclinical mumps is not uncommon, so vaccine will at times be given to some who actually are immune. A single inoculation is adequate. Side effects are minimal and mild.

Polio Vaccine

Routine polio vaccination of adults residing in the US is not recommended (29). Most adults are already immune to polio, the risk of exposure is very small, and there is a slightly greater risk of vaccine-associated paralysis than in children. However, polio vaccination is recommended for susceptible adults who will be at increased risk to exposure by virtue of traveling to an area where polio is common. Because it is not practical to conduct serological tests to determine immunity to polio in prospective travelers, the recommended primary series of polio inoculations should be completed based on past immunization history. The primary series consists of three doses of trivalent oral polio vaccine (TOPV). The first two doses should be given not less than six weeks apart, and the third dose should follow in 8–12 months. Obviously, this recommendation is frequently impractical or there is confusion about the past polio inoculation history. In such instances, at least a single TOPV should be given. A similar "booster" dose should be given to adults who completed their primary series in the distant past.

Yellow Fever Vaccine

The recommended vaccine is the 17-D strain; it should be given to adults traveling to endemic areas. A single dose is given, with a booster dose every ten years.

Rabies Vaccine

Rabies vaccine (30, 31) is administered to adults under two circumstances: as preexposure prophylaxis for laboratory workers and veterinarians who are at unusual risk of exposure to the virus, and as post-exposure prophylaxis. The duck embryo vaccine is currently employed in the US. Local and systemic reactions are frequent, but, in post-exposure prophylaxis, should not contraindicate completion of treatment. In the near future, a new rabies vaccine, grown in human WI-38 cells, will become available. (32) This vaccine requires fewer inoculations and is associated with fewer side reactions.

When rabies exposure has been determined, vaccine is given alone or in combination with human rabies immune globulin. When given alone, 14 daily injections are administered. When given with serum, 21 doses are recommended. These may be given as 21 daily doses, or as two daily doses for the first seven days, and one daily dose the next seven days. Booster doses are given 10 and 20 days after completion of the primary series.

Preexposure prophylaxis consists of either two doses given one month apart followed by a third dose in six months, or three doses five days apart followed by a fourth dose a month later. Serum antibody should be measured after completion of either series, and, if antibody is absent, booster doses given monthly until antibody is present.

BACTERIAL VACCINES

Pneumococcal Vaccine

Despite the availability of effective antibiotics for treatment, *Streptococcus pneumoniae* (pneumococcus) disease is still responsible for considerable morbidity and mortality in the US (33). Pneumococcal pneumonia has an overall mortality of 5–10%, and is a particularly serious disease in the elderly and the debilitated. Pneumococcal meningitis and septicemia remain serious problems. Persons who have undergone splenectomies or have functional asplenia are particularly vulnerable to pneumococcal infection.

A polyvalent capsular polysaccharide vaccine against pneumococcal infections became available in the mid 1970s (33, 34). The vaccine contains capsular polysaccharide from 14 strains of pneumococcus, which accounts for at least 80% of all bacteremic pneumococcal disease in the US. Early studies have shown that vaccine-induced antibody remains at satisfactory levels for at least several years, has been 80% effective in reducing the incidence of bacteremic pneumococcal pneumonia, and is safe.

Definite recommendations for use of the pneumococcal vaccine have not as yet been determined, but the vaccine should logically be given to the elderly, patients with chronic pulmonary and cardiac disease, debilitated persons, asplenic (anatomical and functional) patients, diabetics, and renal patients. A single inoculation is used. Needs and schedules for boosters have yet to be determined. Side effects are common but minor. In some studies, Arthus-type reactions were noted when boosters were given (34).

Meningococcal Vaccines

Three polysaccharide vaccines against diseases caused by *Neisseria meningitides* are available in the US (35, 36). These are monovalent A, monovalent C, and bivalent A-C. Routine use of these vaccines in civilian adult populations are not recommended. The routine use of Type C vaccine has virtually eliminated Type C meningococcus meningitis among military personnel. The vaccines may be given prophylactically to travelers to endemic areas of meningococcal disease, possibly in the control of local epidemics, and as adjunct to antibiotic prophylaxis to household contacts. No vaccine exists for Type B meningococcal disease, which continues to cause the majority of meningococcal disease in the US.

Diphtheria Toxoid

Need for diphtheria immunization in adults is uncommon, although the disease occurs in adults and is usually severe. Rigid surveillance among specific adult populations (i.e. illegal aliens) would undoubtedly magnify the need for adult immunization. Primary immunization requires three doses of combined diphtheria and tetanus toxoid labeled Adult Type (Td). Periods of 4–6 weeks and 6–12 months separate the first and second doses and the second and third doses, respectively.

Tetanus Toxoid

The bulk of the use of tetanus toxoid in adults is in the form of "booster" inoculation given as part of wound management. However, there are adults who have never undergone primary immunization. In adults, tetanus toxoid can be given separately or in combination with diphtheria toxoid (Td). Primary immunization consists of three doses: the second dose 4–6 weeks after the first dose, and the third dose 6–12 months after the second. Boosters in wound management are recommended if five years have elapsed since the completion of the series on a previous booster. In the absence of wound management a booster should be given every 10 years. There is a tendency by the medical profession to "over-booster" for tetanus.

Tetanus vaccination is especially indicated for workers in contact with soil and domestic animals and for persons with increased risk of trauma.

Other Bacterial Vaccines

Cholera, plague, and typhoid vaccines are inactivated vaccines and are primarily indicated for travelers. Typhoid vaccine may also be given to persons with intimate exposure to a known carrier, and to those exposed in institutional outbreaks.

RICKETTSIAL VACCINES

Rocky Mountain Spotted Fever Vaccine

An inactivated vaccine for Rocky Mountain Spotted Fever is available (37, 38). It is recommended for laboratory personnel and individuals at high risk of occupational exposure. In addition, some physicians administer the vaccine to residents of high endemic areas. However, the vaccine is not considered effective, and accurate diagnostic tests for the disease and effective antibiotic treatment exist. Therefore, use of the vaccine under any conditions is questionable.

VACCINES UNDER DEVELOPMENT

Hepatitis B

Several hepatitis B vaccines are currently being tested in humans (39, 40). The vaccines are inactivated and contain only surface antigen. They have proven to be relatively poor antigens in terms of size and persistence of antibody response. They will be indicated for high risk individuals such as renal dialysis patients and personnel, hospital professional and laboratory personnel, and male homosexuals.

Varicella-Zoster

Several experimental varicella-zoster vaccines are under development and have been administered to humans (41). Preliminary results are favorable. Adult varicella pneumonia is a serious disease with a high mortality rate. The availability of an effective vaccine to administer to the suspected susceptible parents of a child with chicken pox would be most welcome.

Cytomegalovirus

Cytomegalovirus infection is a major threat in immunosuppressed persons. Early reports on a live, attenuated cytomegalovirus vaccine are promising (42).

CONCLUDING REMARKS

The use of vaccines in adults is increasing with the development of new vaccines, the changing epidemiology of some common viral diseases, and the continuing lack of specific therapeutic agents. Nonpediatric medical personnel should become familiar with the use of these vaccines.

Literature Cited

1. Rauh, J. L., Schiff, G. M., Johnson, L. B. 1975. Follow-up studies of rubella vaccinees at adolescence. *J. Pediatr.* 86:138–40
2. Krause, P. J., Cheny, J. D., Deseda-Tous, J., et al. 1979. Epidemic measles in young adults. *Ann. Intern. Med.* 90:873–76
3. Melnick, J. L. 1978. Viral vaccines: new problems and prospects. *Hosp. Pract.* 13:104–12
4. Davenport, F. M. 1976. Influenza virus. In *Viral Infections of Humans: Epidemiology and Control,* ed. A. S. Evans. New York: Plenum. 278 pp.
5. Hilleman, M. R. 1977. Serologic responses to split and whole Swine influenza virus vaccines in light of the next influenza pandemic. *J. Infect. Dis.* 136: (Suppl.) S683–S685
6. Barry, D. W., Mayner, R. E., Staton, E., et al. 1976. Comparative trial of influenza vaccines. *Am. J. Epidemiol.* 104:34–46
7. Miller, L. W., Togo, Y., Hornick, R. B. 1975. Clinical and serologic effects of Alice strain live attenuated influenza A (H_3N_2) virus vaccine in an adult population. *Med. Microbiol. Immunol.* 162: 15–21
8. Morris, C. A., Freestone, D. S., Stealey, V. M., Oliver, P. R. 1975. Recombinent WRL 105 strain live attenuated influenza vaccine. *Lancet* 2:196–99
9. Douglas, R. G., Bentley, D. W., Betts, R. F., et al. 1976. Evaluation of a temperature-sensitive influenza virus in elderly and chronically ill subjects. *Am. Rev. Resp. Dis.* 113:293–300
10. Schiff, G. M., Linnemann, C. C. Jr., Shea, L., et al. 1975. Evaluation of a live, attenuated recombinant influenza vaccine in high school children. *Infect. Immun.* 11:754–57
11. Galbraith, A. W., Oxford, J. S., Schild, G. C., et al. 1969. Protective effect of 1-adamantanamine hydrochloride on influenza A-2 infections in the family environment: A controlled double-blind study. *Lancet* 2:1026–28
12. Monto, A. S., Gunn, R. A., Bandyk, M. G., King, C. L. 1979. Prevention of Russian influenza by amantadine. *J. Am. Med. Assoc.* 241:1003–7
13. LaMontagne, J. R., Galasso, G. J. 1978. Report of a workshop on clinical studies of the efficacy of amantadine and rimantadine against influenza virus. *J. Infect. Dis.* 138:928–31
14. Waldman, R. H., Khakoo, R. A., Watson, G. 1978. Isoprinosine: Efficacy against influenza challenge infection in humans. *Curr. Chemother.* 1:368
15. Betts, R. F., Douglas, R. G. Jr., George, S. D., Rinehart C. J. 1978. Isoprinosine

in experimental influenza A infections in volunteers. *Abstr. 78th Ann. ASM Meet.,* No. A75
16. Schiff, G. M., Roselle, G., Young, B., et al. 1978. Clinical evaluation in artifically induced influenza in humans. *Abstr. 78th Ann. ASM Meet.,* No. A74
17. Horstmann, D. M. 1975. Controlling rubella: problems and perspectives. *Ann. Intern. Med.* 83:412–17
18. Schiff, G. M., Rauh, J. L., Rotte, T. 1969. Rubella vaccine evaluation in a public school system. *Am. J. Dis. Child.* 118:203–8
19. Plotkin, S., Farquhar, J. D., Ogra, P. L. 1973. Immunologic properties of RA 27/3 rubella virus vaccine. A comparison with strains presently licensed in the United States. *J. Am. Med. Assoc.* 225:585–90
20. Modlin, J. F., Herrmann, K., Brandling-Bennett, A. D., et al. 1976. Risk of congenital abnormality after inadvertant rubella vaccination of pregnant women. *N. Engl. J. Med.* 294(18): 972–74
21. Schiff, G. M., Rauh, J. L., Young, B., et al. 1978. Rubella-vaccinated students: follow-up in a public school system. *J. Am. Med. Assoc.* 240:2635–37
22. Schiff, G. M., Linnemann, C. C. Jr., Shea, L., et al. 1974. Rubella surveillance and immunization among college women. *Obstet. Gynecol.* 43:143–47
23. Schiff, G. M., Rauh, J. L. 1971. Rubella control. *Am. J. Dis. Child.* 122:112–16
24. Center for Disease Control. 1978. Goal to eliminate measles from the United States. *Morbid. Mortal. Week. Rep.* 27:391
25. Krugman, S. 1977. Present status of measles and rubella immunization in the United States: a medical progress report. *J. Pediatr.* 90:1–12
26. Rand, K. H., Emmons, R. W., Merigan, T. C. 1976. Measles in adults. An unforeseen consequence of immunization? *J. Am. Med. Assoc.* 236:1028–31
27. Martin D. B., Werner, L. B., Nieburg, P. I., Blair, D. C. 1979. Atypical measles in adolescents and young adults. *Ann. Intern. Med.* 90:877–81
28. Lewy, R. M. 1977. Preventing mumps and its complications. *Am. Family Physician* 16:119–21
29. Nightengale, E. O. 1977. Recommendations for a national policy on poliomyelitis vaccination. *N. Engl. J. Med.* 297:249–53
30. Campbell, J. B., Kaplan, M. M., Koprowski, H., et al. 1968. Present trends and the future in rabies research. *Bull. World Health Org.* 38:373–81
31. Turner, G. S. 1977. An assessment of the current position of rabies vaccination in man. In *Recent Advances in Clinical Virology,* ed. A. P. Waterson, pp. 79–92. London: Churchill Livingston
32. Kuwert, E. K., Marcus, I., Werner, J., et al. 1977. Post-exposure use of human diploid cell culture rabies vaccine. *Develop. Biol. Stand.* 37:273–86
33. Austrian, R. 1976. Vaccines of pneumococcal capsular polysaccharides and the prevention of pneumococcal pneumonia. *The Role of Immunological Factors in Infectious, Allergic, and Autoimmune Processes,* ed. R. F. Beers Jr., E. G. Barret, pp. 79–89. New York: Raven
34. Hilleman, M. R., McLean, A. A., Vella, P. P., et al. 1978. Polyvalent pneumococcal polysaccharide vaccines. *Bull. World Health. Org.* 56:371–75
35. Artestein, M. S. 1975. Control of meningococcal meningitis with Meningococcal vaccines. *Yale J. Biol. Med.* 48:197–200
36. Mäkelä, P. H., Kāyhty, H., Weckström, P., et al. 1975. Effect of group A meningococcal vaccine in army recruits in Finland. *Lancet* 2:883–86
37. Kenyon, R. H., Sammons, L. S. C., Pedersen, C. E. Jr. 1975. Comparisons of three Rocky Mountain spotted fever vaccines. *J. Clin. Microbiol.* 2:300–4
38. Ascher, M. S., Oster, C. N., Harber, P. L., et al. 1978. Initial clinical evaluation of a new Rocky Mountain spotted fever vaccine of tissue culture origin. *J. Infect. Dis.* 138:217–21
39. Maupas, P., Goudeau, A., Coursaget, P., et al. 1978. Hepatitis B vaccine: efficacy in high-risk settings, a two-year study. *Intervirology* 10:196–208
40. Hilleman, M. R., Buynak, E. B., Roehm, R. R., et al. 1975. Purified and inactivated human hepatitis B vaccine: progress report. *Am. J. Med. Sci.* 270:401–4
41. Kempe, C. H. 1977. Varicella vaccine at the crossroads. *Pediatrics* 60:930–31
42. Plotkin, S. 1976. CMV vaccine prepared in WI-38. *Dev. Biol. Stand.* 37:301–5

Ann. Rev. Med. 1980. 31:453–62
Copyright © 1980 by Annual Reviews Inc. All rights reserved

COMMUNITY CARDIOPULMONARY RESUSCITATION

❖7373

Leonard A. Cobb, M.D., Alfred P. Hallstrom, Ph.D.,
Robert G. Thompson, M.D., Lynn P. Mandel, M. A.
and Michael K. Copass, M.D.[1]

Department of Medicine, Department of Biostatistics, and Office of Research in Medical Education, University of Washington and Harborview Medical Center, Seattle, Washington 98104

The majority of deaths due to coronary heart disease are sudden, unexpected events that occur outside the hospital (1). This represents a formidable challenge since treatment aimed at the problem of sudden cardiac death must begin in the community and must be initiated rapidly.

Most instances of sudden cardiac death are the result of ventricular fibrillation and are usually due to underlying coronary heart disease (2). However, only a minority of these episodes, some 20–40%, appear to be manifestations of acute myocardial infarction; the majority are primary dysrhythmic events (3, 4). Ventricular fibrillation per se can usually be reverted by the prompt application of a transthoracic dc shock (5, 6). The major obstacle to the successful management of cardiac arrest is the inherent delay in the initiation of resuscitation outside the hospital. Unless the circulation of oxygenated blood is reestablished within approximately four minutes, irreversible anoxic cerebral injury will ensue.

CARDIOPULMONARY RESUSCITATION

Methods to revive apparently dead persons have undergone centuries of trials and revisions. The development of mouth-to-mouth ventilation (7)

---

[1] Supported by grants R18-HS-01943 from the National Center for Health Services Research and HL18805 from the National Heart Lung and Blood Institute.

0066-4219/80/0401-0453$01.00

was an important milestone in modern resuscitation. However, it was not until 1961, when Jude, Kouwenhoven & Knickerbocker (8) reported the application of external chest compression, that effective circulatory support became a reality. Closed chest compression and positive pressure ventilation are now widely recognized as effective measures to provide circulation and ventilation to the patient with cardiac arrest (9). Proper application of these measures is referred to as cardiopulmonary resuscitation (CPR) or basic life support:

> "Basic life support is an emergency first aid procedure that consists of recognizing respiratory and cardiac arrest and starting the proper application of cardiopulmonary resuscitation to maintain life until a victim recovers sufficiently to be transported or until advanced life support is available. This includes the A-B-C steps of cardiopulmonary resuscitation: a. airway, b. breathing, c. circulation" (9).

OUT-OF-HOSPITAL ADVANCED LIFE-SUPPORT

For the victim of a cardiac arrest due to ventricular fibrillation, CPR is a temporizing measure to deliver oxygen to the brain and other tissues until definitive therapy (advanced life-support) is available. Electrical defibrillation is always necessary; in addition, effective therapy usually includes the administration of intravenous drugs and provision of airway protection, preferably tracheal intubation.

Advanced level, prehospital emergency care systems were developed approximately 10 years ago in the United States (10–13). Most of these systems provide comprehensive emergency care for all out-of-hospital medical and surgical emergencies.

The Seattle System

The Seattle Fire Department's emergency care system (Medic I) represents a tiered response system that provides basic and advanced life-support. Appropriate units are dispatched according to medical considerations and the site of the incident. This emergency care system also incorporates important participation by the lay public in the early application of CPR (11).

The initial response to most medical emergencies is from a nearby fire station, with either a fire engine or an ambulance type of vehicle, both manned by emergency medical technicians trained in basic first-aid and CPR. The average response time from dispatch until arrival on the scene is slightly less than three minutes. When there are life-threatening emergencies, e.g. an unconscious patient or one with symptoms of a heart attack, an advanced life-support unit is simultaneously dispatched. The system is schematically represented in Figure 1.

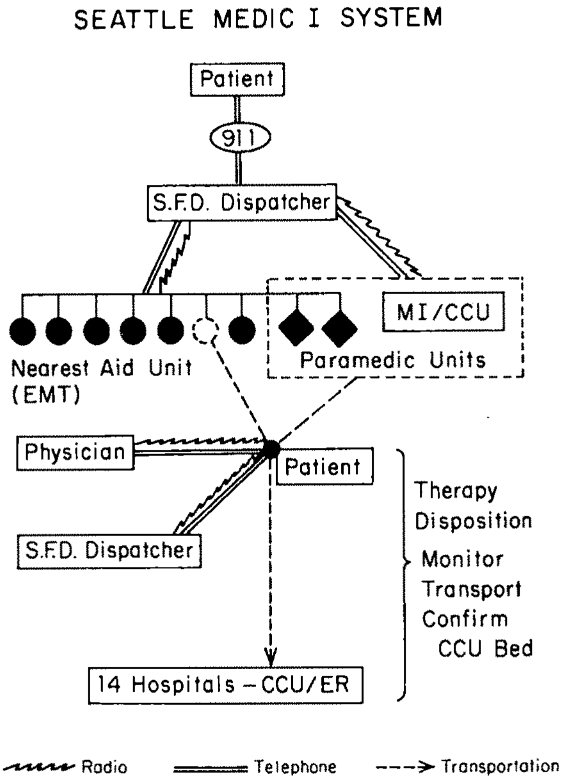


Figure 1 Schema for the Seattle Fire Department's (S.F.D.) emergency care system. Initial response is provided by aid units (or by fire engine companies, not depicted here). In life-threatening situations a mobile intensive/coronary care unit (MI/CCU) is dispatched simultaneously with an aid unit. CCU = coronary care unit; ER = hospital emergency room.

The secondary (advanced life-support) units are staffed by paramedical personnel who have received approximately 1600 hours of training; they are skilled and knowledgeable in most facets of emergency care, including tracheal intubation, arrhythmia monitoring and detection, use of a defibrillator, and the administration of common emergency drugs. Standing, written orders are utilized for the management of cardiac arrest and hypovolemic shock. For all other conditions, therapy is carried out in conjunction with a physician through radio or telephone communication. The goal of this program is to deliver out-of-hospital emergency care comparable to that which a well-trained physician would provide at the scene.

EXPERIENCE TREATING PATIENTS WITH VENTRICULAR FIBRILLATION Each year approximately 300 patients in Seattle are treated for ventricular fibrillation present on arrival of paramedical personnel. An additional 40 to 50 patients develop ventricular fibrillation and cardiac arrest while under surveillance of the paramedics. In the vast majority of cases of ventricular fibrillation, the request for medical aid is because of sudden loss of consciousness—most often without premonitory symptoms.

At present, approximately 60% of patients with ventricular fibrillation (VF) are resuscitated with restoration of spontaneous pulse and measurable blood pressure, and approximately 30% of VF patients are discharged home following resuscitation (14). We have found that survival in these patients is associated with rapid fire-department response times (15) and with participation of bystanders in initiating CPR.

Participation of bystanders in CPR Because someone other than the victim of cardiac arrest initiates the call for help, nearly all cases treated outside the hospital have the potential for early initiation of CPR by bystanders. Nevertheless, in 1970–1971 bystanders initiated CPR in only about 5% of cases treated in Seattle (15). By 1976, this number had increased to 34%. In the suburban areas of King County adjacent to Seattle, 22% of resuscitation efforts were reported to be initiated by bystanders prior to arrival of emergency medical technicians (22). Much of this increase was certainly due to the fire department and other community training programs described below.

Relationship of bystander-initiated CPR to mortality Table 1 shows the improved survival reported for bystander-initiated CPR in patients treated for cardiac arrest outside the hospital. It is noteworthy that survival in patients treated promptly for out-of-hospital cardiac arrest appears to be greater than that reported for patients treated for cardiac arrest in the hospital, where overall survival to hospital discharge averages about 16% (16). This is probably because cardiac arrest in ambulatory patients is usually a "primary" rhythm disturbance, whereas cardiac arrest in hospital-

Table 1 Relationship between bystander-initiated CPR and survival to hospital discharge

				Survival		
Community (Ref.)	Cases (no.)	Average response times (min)[a]	Type of cardiac arrest	With bystander CPR	Without bystander CPR	Probability[b]
Oslo (17)	631	not stated	not stated	27/75 (36%)	43/556 (8%)	<.001
Surburban Seattle (22)	487	4.2	all cardiac	25/108 (23%)	45/379 (12%)	<.01
Seattle (14)	316	2.9	VF on arrival[c]	47/109 (43%)	43/207 (21%)	<.001

[a] Average time after dispatch of the first responding emergency vehicle until arrival on the scene.
[b] Probability (by Chi square) of chance difference in survival rates with and without bystander CPR.
[c] VF on arrival = ventricular fibrillation present when paramedical personnel arrived.

ized patients is often secondary to, or associated with, advanced symptomatic illness.

Reduction of neurologic sequelae As might be expected, more rapid application of CPR results in substantially fewer neurologic complications of anoxia. In the Seattle report (14), hospital deaths due to anoxic coma were fewer, and rapid neurologic recovery was usually observed when bystanders promptly initiated CPR. The proportion of resuscitated patients who regained consciousness in the first 24 hours following resuscitation was greatly increased (61% compared to 9%), and the incidence of prolonged coma and persistent disorientation were both significantly reduced in patients who received bystander-initiated CPR compared to those who did not (14) (Figure 2). Lund & Skulberg (17) also noted less brain damage when CPR was initiated by lay persons prior to arrival of ambulance personnel. These findings are in accord with the report by Copley and colleagues (18), who described 12 patients hospitalized after resuscitation from cardiac arrest; these authors noted improved survival and less severe neurologic sequelae in patients who had received CPR from bystanders. In the Copley study, there was also less myocardial dysfunction as evidenced by lower pulmonary wedge pressures and higher cardiac outputs in patients who had received bystander-initiated CPR.

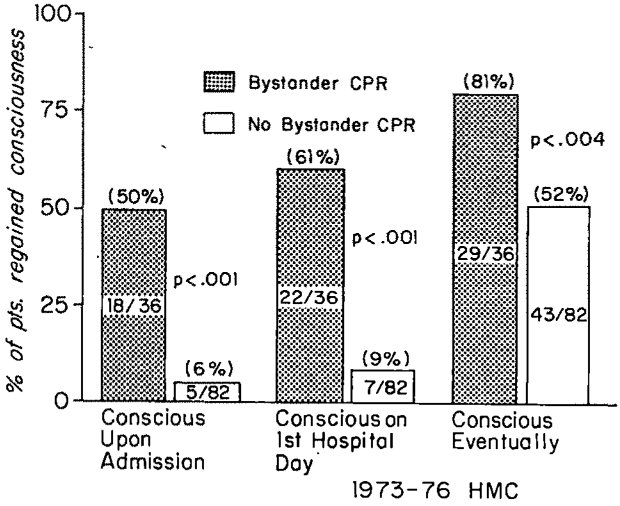


Figure 2 Time until return of consciousness in 118 patients admitted consecutively to Harborview Medical Center (HMC) after resuscitation from out-of-hospital ventricular fibrillation. The proportion who regained consciousness was greater and the period of unconsciousness was shorter in cases where bystanders had initiated CPR. (Reproduced with permission of *Annals of Internal Medicine*.)

DEVELOPMENT OF A COMMUNITY CPR PROGRAM

Seattle's community CPR program (Medic II) was launched in the Fall of 1971 as a direct consequence of an informal discussion relating to response times and the participation of fire-combat personnel in emergency medical care. Gordon F. Vickery, then Chief of the Seattle Fire Department, suggested offering CPR training to a large segment of the general public: If basic CPR skills were appropriate for all firefighters to know and if these skills could be learned relatively easily, why not make such training available to the general public?

The rationale for the community CPR training program was straightforward: In instances of cardiac arrest in which the emergency care system was involved, bystanders were virtually always on the scene prior to the arrival of fire-department personnel (someone other than the patient had requested assistance). If such bystanders were able and willing to initiate CPR, the chances of surviving a cardiac arrest might well be improved. The initial goal of the Medic II program was to train 100,000 persons, about one fifth of Seattle's population. To facilitate public awareness of the new program, substantial support was mobilized through the news media and several service organizations.

Course Description

The Medic II CPR training program is offered without charge to groups and individuals requesting such training. The classes are held at a location and time convenient for the students and are taught by specially trained Seattle firefighter/paramedics. Classes are usually given as a single three-hour session, typically in the evening, although classes for high school students may be separated into three one-hour sessions. The material is first presented in a lecture-demonstration of approximately 45 minutes, based on the recommendations of the National Academy of Sciences/American Heart Association made in 1966 and subsequently revised in 1974 (9). Students are provided information about (*a*) cardiac arrest and the appropriate diagnostic measures to be used in its recognition, (*b*) the principles and techniques in application of CPR, (*c*) the symptoms of a heart attack, (*d*) whom to call for assistance (the Seattle Fire Department) and how to call (dial 911). A CPR training film ("Pulse of Life"; later "New Pulse of Life") is shown, followed by a short recapitulation with opportunities for questions and answers. All students are given the opportunity for supervised CPR practice on a life-like training manikin. In addition, there is an opportunity at the end of the session for discussion of other medical emergencies and a description of the interaction of the Fire Department's Medic I program with hospitals and physicians in the community. The students are expected to show proficiency in CPR, but a formal certification process is not attempted. Trainees receive a card indicating that they have attended a basic CPR training class approved by the American Heart Association of Washington.

EVALUATION OF THE CPR TRAINING PROGRAM

Evaluation has been carried out in several phases. First, we conducted a brief demographic assessment of the students and obtained an overall impression of their evaluation of the course. Second, testing was done immediately after the course was completed to assess CPR knowledge and skills. Finally, repeat testing was done on samples of students at six months, one year, and two years after completion of the course to determine skills retention.

The Students

As of August, 1979, over 223,000 students had been trained in the Seattle Fire Department's Medic II CPR training program, approximately 150,000 in the city and 73,000 in the suburban communities adjacent to Seattle.

During 1976, 1700 randomly selected students completed a questionnaire at the time of the CPR class. On the basis of this questionnaire, the students were judged to be young (average 33 years), well educated (70% had some education beyond high school), and mostly white-collar workers. Ten percent of the students had taken the Medic II class previously, and 15% had received some other form of CPR training at an earlier date. Only one third of the responders resided in households in which there was a person over age 50. Almost all students felt the class presented useful information, and nearly 90% indicated their willingness to take a refresher course.

Testing Immediately After Class

Immediately after training, 151 randomly selected students from 17 classes were tested on a recording manikin. The students were graded on a total of 37 variables obtained from analysis of the manikin tape print-outs and the ratings made by a trained nurse-observer. The American Heart Association's Basic Cardiac Life Support Standards (19) were revised slightly to define a basic level of competence or performance judged adequate to keep a victim alive until fire-department personnel arrived. Performance that met the American Heart Association's standards was recorded as excellent; meeting the revised criteria was considered satisfactory; performance below this level was considered unsatisfactory. Overall, performance on most CPR skills was judged to be at least satisfactory in 70% or more of the students. Satisfactory depth of compression, compression rate, and ventilation rates were achieved in 73–99% of the tests immediately after the classes. However, satisfactory performance of certain isolated skills, e.g. adequate ventilation volume and checking of the chest expansion during ventilation, was achieved less often.

Retention Testing

Forty-three of the subjects were tested immediately after class completion and again six months later. The six-month scores were uniformly lower than the initial scores. However, of 13 skills considered the most important for effective CPR, only three showed significant declines ($P < .05$): initial four ventilations, checking the carotid pulse, and proper hand position (20). In spite of this, performance still remained satisfactory for at least 50% of the students on most skills.

After one year, 52 students were retested. Of these, 33 had also been tested at six months while 19 had only initial and one-year testing. The subjects tested at one year had consistently better retention scores if they had been previously tested at six months. The ten-minute practice and feedback of the first retesting presumably acted as a refresher that facilitated CPR performance one year later. The same trend was seen on two-year retention where students tested initially and only at two years showed a sharp decline in competency.

We have demonstrated that most lay citizens can perform reasonably adequate CPR on a recording manikin immediately after training. As shown by our students, as well as by those described by Weaver and colleagues (21), knowledge and skills decline after training, indicating a need for development of simple mechanisms to reinforce CPR skills.

SUMMARY, CONCLUSIONS, AND QUESTIONS

It is our conclusion that the availability of bystanders ready and able to initiate CPR represents an effective adjunct to a rapid-response, emergency medical care system. It is important to emphasize the adjunctive nature of citizen CPR participation; for the majority of patients who require CPR, basic life-support is a temporizing measure until definitive therapy in the form of defibrillation, airway protection, and medications can be instituted.

Clearly the average person may have the opportunity to perform CPR perhaps only once or twice in his lifetime. Hence, it is impractical to expect the lay public to be as expert in CPR as health care professionals who use these techniques regularly. However, we believe that citizens should be taught with the expectation of their being able to perform CPR in a reasonably proficient manner.

Questions

Several questions commonly arise when considering the provision of CPR instruction to the general public.

DOES THE QUALITY OF LAY CPR PERFORMANCE AFFECT THE PATIENT'S OUTCOME? In interviewing bystanders, we have been unable to detect an obvious improvement in patient outcome related to the extent of the rescuer's formal training or the apparent quality of CPR as described. The person with some training in CPR also has an awareness of the problem of cardiac arrest, should know whom and when to call for help, and is likely to be aware of the importance of avoiding delays. Under the conditions of a rapid-response system where the time from dispatch until arrival on the scene approximates two to four minutes, the quality of bystander-initiated CPR is probably less important than the fact that he or she is initiating CPR and that skilled help is on the way. It is quite possible that even suboptimal attempts at CPR during the minutes before arrival of the paramedics might forestall the onset of anoxic cerebral injury. However, if the bystander were required to carry out CPR for more than a few minutes, it is likely that the quality of the resuscitation effort would affect survival considerably more.

WHO SHOULD BE TRAINED? It is our feeling that virtually everyone should be taught when and how to perform cardiopulmonary resuscitation. CPR training is being introduced into school curricula, and it is likely that this practice will be enhanced. In the meanwhile, it would seem prudent for the families of persons with known coronary heart disease to have this training. Most cardiac arrests occur in the home where the only available rescuer is often the patient's spouse. Physicians' offices and hospital coronary care units would appear to be excellent sites to initiate referral for CPR instruction of spouses and other family members.

ARE THERE MEDICAL OR LEGAL RISKS? A major concern is that CPR might be applied to a person who does not have circulatory arrest and that unnecessary damage could ensue. This is a possible but probably an unlikely happening. The person who faints will deter his would-be-rescuer as soon as consciousness returns. For the patient who has had a cardiac arrest, it is difficult to envision a more damaging situation than not making any attempt to carry out CPR. Most states have so-called Good Samaritan laws to provide immunity from civil liability for rescuers acting in good faith. In addition, an extensive search of medical, legal, and lay literature, personal inquiries, and solicitation of nationwide interest groups failed to provide a single case in which liability was found for a layman's performance of CPR (23).

ARE THERE NEEDS FOR REFRESHER TRAINING? There is an appreciable decline in CPR skills and knowledge over the course of time. As mentioned above, it appears impractical to expect expert or near-expert CPR performance from the general public. However, we think it reasonable to develop and test the use of simple, short refresher devices such as illustrated printed materials, televised instruction, or manikin practice with feedback capability, or perhaps combinations of these.

Literature Cited

1. Gordon, T. 1971. Premature mortality from coronary heart disease. The Framingham study. *J. Am. Med. Assoc.* 215:1617–25
2. Lawn, B. 1979. Sudden cardiac death: The major challenge confronting contemporary cardiology. *Am. J. Cardiol.* 43:313–28
3. Baum, R. S., Alvarez, H. III, Cobb, L. A. 1974. Survival after resuscitation from out-of-hospital ventricular fibrillation. *Circulation* 50:1231–35
4. Liberthson, R. R., Nagel, E. L., Hirschman, J. C., Nussenfield, J. D., Blackbourne, B. D., Davis, J. D. 1974. Pathophysiologic observations in prehospital ventricular fibrillation and sudden cardiac death. *Circulation* 49:790–98
5. Campbell, N. P. S., Webb, S. W., Adgey, A. A. J., Pantridge, J. F. 1977. Transthoracic ventricular defibrillation in adults. *Br. Med. J.* 2:1379–81
6. Crampton, R. S., Gascho, J. A., Cherwek, M. L., Sipes, J. N. 1978. Low energy and fast serial DC shock ventricular defibrillation in man. *Med. Instrum. Baltimore* 12:53–55
7. Elam, J. O., Greene, D. G. 1961. Mission accomplished: Successful mouth-to-mouth resuscitation. *Anesth. Analg. Cleveland* 40:578–80
8. Jude, J. R., Kouwenhoven, W. B., Knickerbocker, G. G. 1961. Cardiac arrest: Report of application of external cardiac massage on 118 patients. *J. Am. Med. Assoc.* 178:1063–70
9. Standards for cardiopulmonary resuscitation (CPR) and emergency cardiac care (EEC) 1974. *J. Am. Med. Assoc.* 174:834–68
10. Nagal, E. L., Hirschman, J. C., Nussenfeld, S. R., Rankin, D., Lunblad, E. 1970. Telemetry-medical command in coronary and other mobile emergency care systems. *J. Am. Med. Assoc.* 214:332–38
11. Cobb, L. A., Alvarez H., Copass, M. K. 1976. A rapid response system for out-of-hospital emergencies. *Med. Clin. North Am.* 60:283–90
12. Lewis, R. P., Warren, J. V. 1974. Factors determining mortality in the prehospital phase of acute myocardial infarction. *Am. J. Cardiol.* 33:152 (Abstr.)
13. Lewis, A. J., Criley, J. M. 1974. An integrated approach to acute coronary care. *Circulation* 50:203–5
14. Thompson, R. G., Hallstrom, A. P., Cobb, L. A. 1979. Bystander-initiated cardiopulmonary resuscitation in the management of ventricular fibrillation. *Ann. Intern. Med.* 90:737–40
15. Alvarez, H., Cobb, L. A. 1975. Experience with CPR training of the general public. In *Proc. Natl. Conf. on Standards for Cardiopulmonary Resuscitation and Emergency Cardiac Care,* pp. 33–37. Dallas: American Heart Association, Inc.
16. Stephenson, H. E. Jr. 1974. *Cardiac Arrest and Resuscitation,* pp. 827–44. St. Louis, Mo: Mosby. 998 pp.
17. Lund, I., Skulberg, A. 1976. Cardiopulmonary resuscitation by lay people. *Lancet* 2:702–5
18. Copley, D. P., Mantle, J. A., Rogers, W. J., Russell, R. O. Jr., Rackley, C. E. 1977. Improved outcome for prehospital cardiopulmonary collapse with resuscitation by bystanders. *Circulation* 56:901–5
19. American Heart Association. 1978. *A Manual for Instruction of Basic Cardiac Life Support.* Dallas, Texas
20. Mandel, L. P., DeMers, J. L., Olsen, M. K., Cobb, L. A. 1977. Evaluation of a citizen CPR training program *Circulation Suppl.* III 55 & 56:114 (Abstr.)
21. Weaver, F. J., Ramirez, A. G., Dorfman, S. B., Raizner, A. E. 1979. Trainees' retention of cardiopulmonary resuscitation—how quickly they forget. *J. Am. Med. Assoc.* 241:901–3
22. Eisenberg, M., Bergner, L., Hallstrom, A. 1979. Paramedic programs and out-of-hospital cardiac arrest: I. Factors associated with successful resuscitation. *Am. J. Publ. Health* 69:30–38
23. McIntyre, K. M. 1976. Status of liability risk for out-of-hospital cardiopulmonary resuscitation. *Circulation Suppl.* II 54:224 (Abstr.)

Ann. Rev. Med. 1980. 31:463–87
Copyright © 1980 by Annual Reviews Inc. All rights reserved

THE NATURAL HISTORY AND RESPONSE TO THERAPY OF LUPUS NEPHRITIS

◆7374

E. M. Ginzler, M.D., A. J. Bollet, M.D., and E. A. Friedman, M.D.

Department of Medicine, State University of New York,
Downstate Medical Center, Brooklyn, New York 11203

Analysis of the natural history and responsiveness to therapy of lupus nephritis has been made possible by the study of percutaneous renal biopsies. Subsets of patients with lupus nephritis can be defined on the basis of histologic classification. For determining options in treatment and for providing a guide to prognosis, each patient with two normal- or large-sized kidneys who has active nephritis (proteinuria, microscopic hematuria, azotemia) should have a renal biopsy. Biopsy specimens are then studied by light, fluorescence, and electron microscopy.

HISTOLOGIC CLASSIFICATION

Two major systems of classification of lupus nephritis are currently in widespread use. One is based upon the histologic appearance of the glomerulus as seen on light microscopy (LM) and is a modification of the classification outlined by Baldwin et al in 1970 (1). The second, proposed by the World Health Organization (WHO), considers both the light microscopic appearance and the localization of immune complexes as seen by electron microscopy (EM) and immunofluorescence (IF) localization of labeled heterologous antihuman antisera (2).

Renal biopsy in a patient with systemic lupus erythematosus may reveal glomeruli that are entirely normal by light microscopy. If there is no

0066-4219/80/0401-0463$01.00

evidence of immune complex deposition by IF or EM, the kidney is termed normal (WHO Class I). More commonly, however, in glomeruli that appear normal by LM, immune complexes confined to the mesangium are found by IF (Figure 2*a*) or EM (Figure 3*a*), and are classified as WHO Class IIA. Mesangial immune complex deposition may also be accompanied by a histologic appearance of increased mesangial matrix and/or areas of mesangial hypercellularity, noted in some or all glomeruli (Class IIB) (Figure 1*a*.) These two degrees of minimal pathology are termed mesangial lupus nephritis.

Baldwin et al's classification of focal proliferative lupus nephritis (FPLN) corresponds to WHO Class III. In addition to the mesangial changes already discussed, FPLN is characterized by capillary endothelial cell proliferation, necrosis, and infiltration with polymorphonuclear leukocytes. Characteristic of Class III is a limitation of the pathologic changes to some lobules (segments) of the glomerulus, while other lobules remain normal on LM (Figure 1*b*). Also typical of Class III is some affected glomeruli and others entirely normal. By WHO definition in Class III, less than 50% of glomeruli are involved. IF and EM examination show abundant mesangial and often subendothelial (Figure 3*b*) immune deposits, which may involve all or a portion of glomeruli.

WHO Class IV, equivalent to Baldwin et al's diffuse proliferative lupus nephritis (DPLN), is characterized by the same qualitative changes on LM as seen in focal nephritis, except that each glomerulus is entirely involved, and the changes are seen throughout all or nearly all glomeruli (Figure 1*c*). Immune deposits are generally intense, and present in a subendothelial as well as mesangial location (Figure 2*b* and 3*b*). A membranoproliferative variant of diffuse nephritis also included in Class IV occurs, in which the predominant histologic change is thickening of the capillary basement membrane, with some proliferative and/or necrotizing changes.

Membranous lupus nephritis, WHO Class V, is characterized by diffusely thickened glomerular capillary walls, without endothelial hypercellularity or necrosis (Figure 1*d*). Immune deposits are present in a subepithelial location along the glomerular capillary basement membrane on EM (Figure 3*c*) and are distributed throughout the glomerulus on IF (Figure 2*c*).

Pathologic changes in the kidney in systemic lupus erythematosus (SLE) are not limited to the glomerulus. Tubular fatty change and/or atrophy is often observed adjacent to glomeruli with proliferative pathology. Infiltration with mononuclear cells may occur throughout the interstitium, either in scattered focal areas or diffusely (3). Pathologic changes may also be present in small arteries and afferent arterioles, with inflammatory infiltration, intimal thickening, and/or fibrinoid necrosis, as is more commonly observed in the kidney in periarteritis nodosa (4).

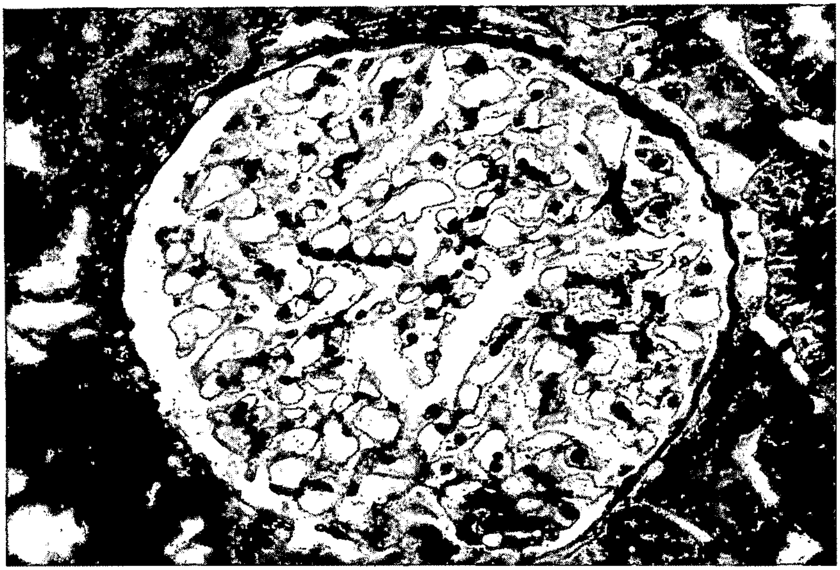


Figure 1a Light microscopic appearance of WHO Class IIB (mesangial) lupus nephritis showing slight mesangial hypercellularity but otherwise normal glomerular capillary loops.

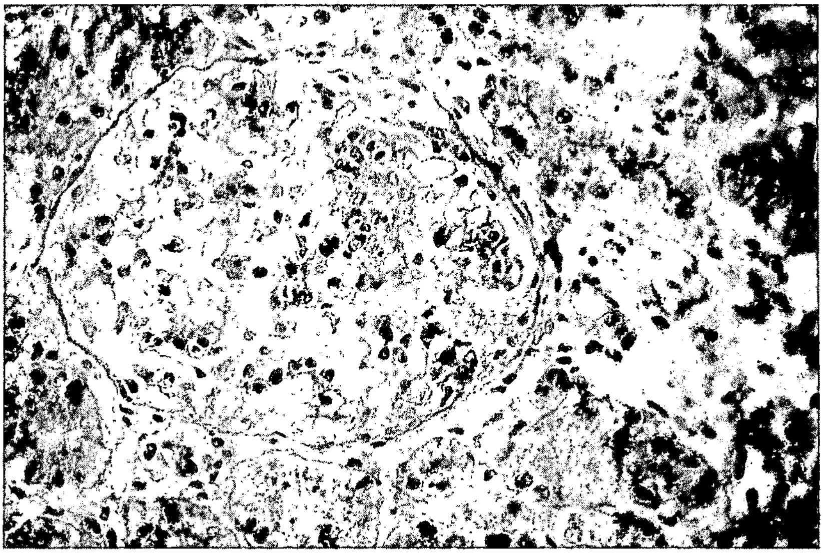


Figure 1b Light microscopic appearance of Class III (focal proliferative) lupus nephritis showing segmental proliferation and basement membrane thickening with other lobules of the glomerulus remaining normal.

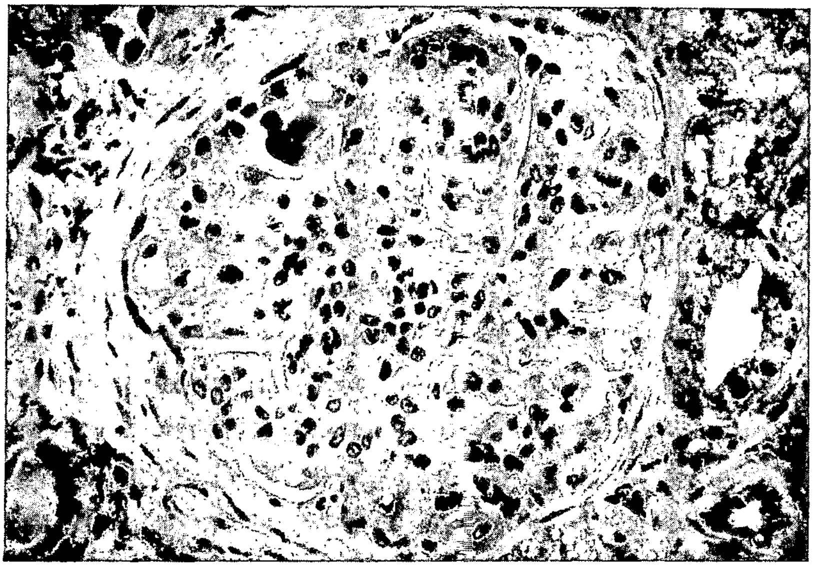


Figure 1c Light microscopic appearance of Class IV (diffuse proliferative) lupus nephritis showing basement membrane thickening and endothelial cell proliferation throughout the entire glomerulus.

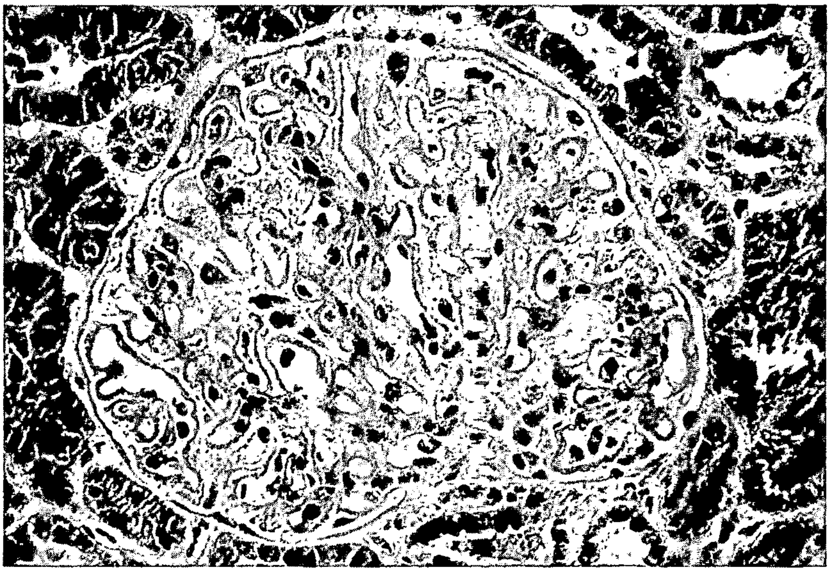


Figure 1d Light microscopic appearance of Class V (membranous) lupus nephritis showing diffuse basement membrane thickening without cellular proliferation or necrosis.

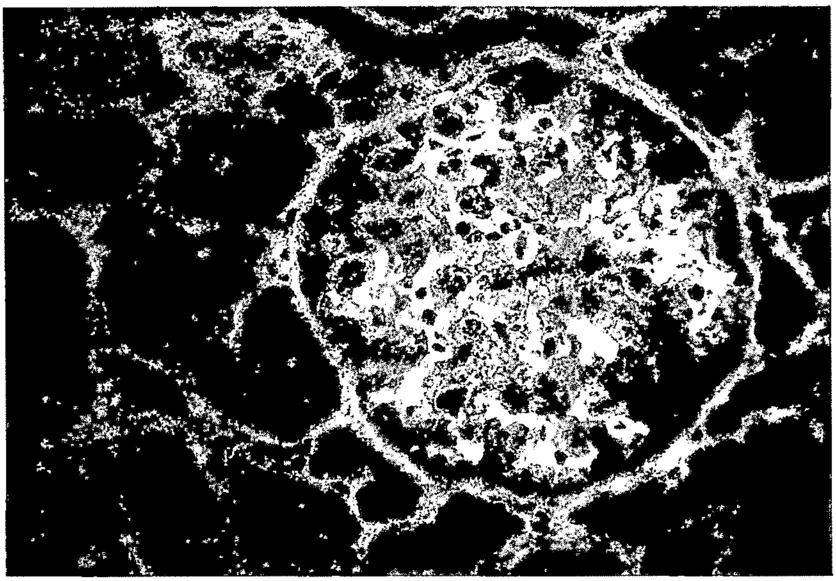


Figure 2a Immunofluorescent staining illustrating deposition of IgG localized to the mesangium in Class II lupus nephritis.

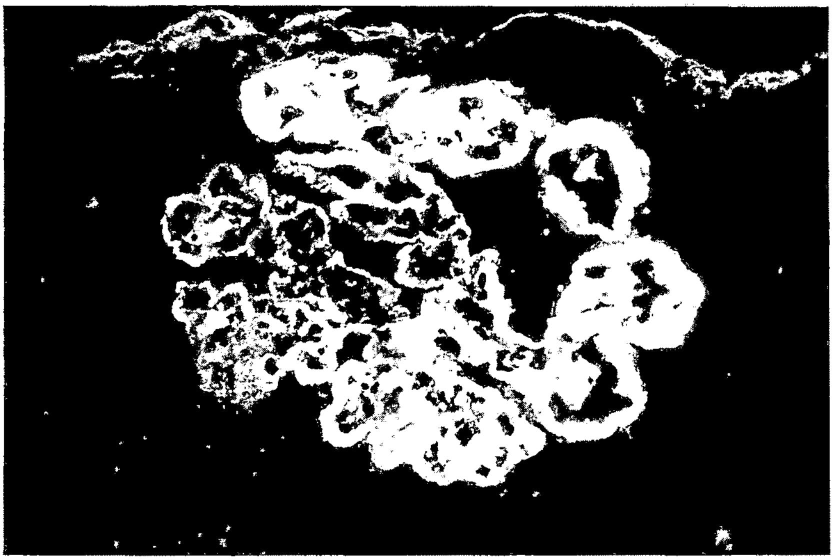


Figure 2b Immunofluorescent staining of immune deposits containing IgG distributed throughout the glomerulus, present here in diffuse proliferative (Class IV) lupus nephritis, but also seen less densely in focal proliferative (Class III) nephritis.

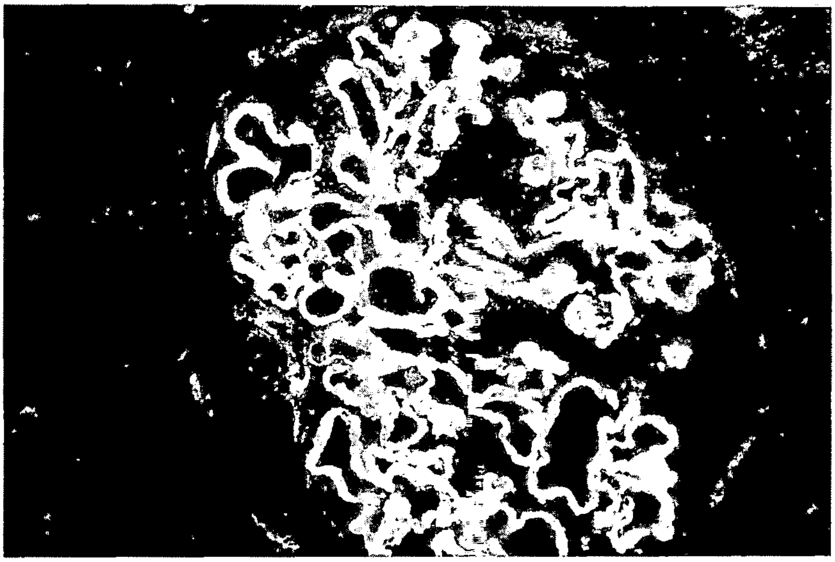


Figure 2c Immunofluorescent staining for C3 showing immune complex deposition within glomeruluar capillary loops in membranous (Class V) lupus nephritis.

CLINICAL CORRELATIONS AND RESPONSE TO THERAPY

Occasionally Class IV or V glomeruli may be discovered in patients with minimal expression of renal disease (5–7). But generally the degree of morphologic change in the kidney correlates with the severity of the patient's clinical illness. Mesangial nephritis is typically associated with absence of clinical abnormalities, as measured by creatinine clearance, urinary sediment, and serologic studies. Indeed, Baldwin et al suggested that probably all patients with SLE have some degree of mesangial immune complex deposition (4). In those with renal pathology (Class IIB) abnormalities are usually limited to proteinuria, usually under 1 g/24 hr, and 5–15 red and/or white blood cells per high power field in the spun urine sediment. Serum complement, measured as C3 or CH_{50}, may be decreased, and anti-DNA antibodies may be present. Endogenous creatinine clearance, an excellent measure of glomerular filtration rate, is almost always normal.

Clinical renal disease is generally present, though mild, in patients with FPLN (Class III). Erythrocytes and red cell casts in the urine sediment, proteinuria in the range of 0.5–2.0 g/24 hr hypocomplementemia, and the presence of antibodies to native DNA are common. The nephrotic syndrome (urinary protein exceeding 3.5 g/24 hr) and hypertension are rare

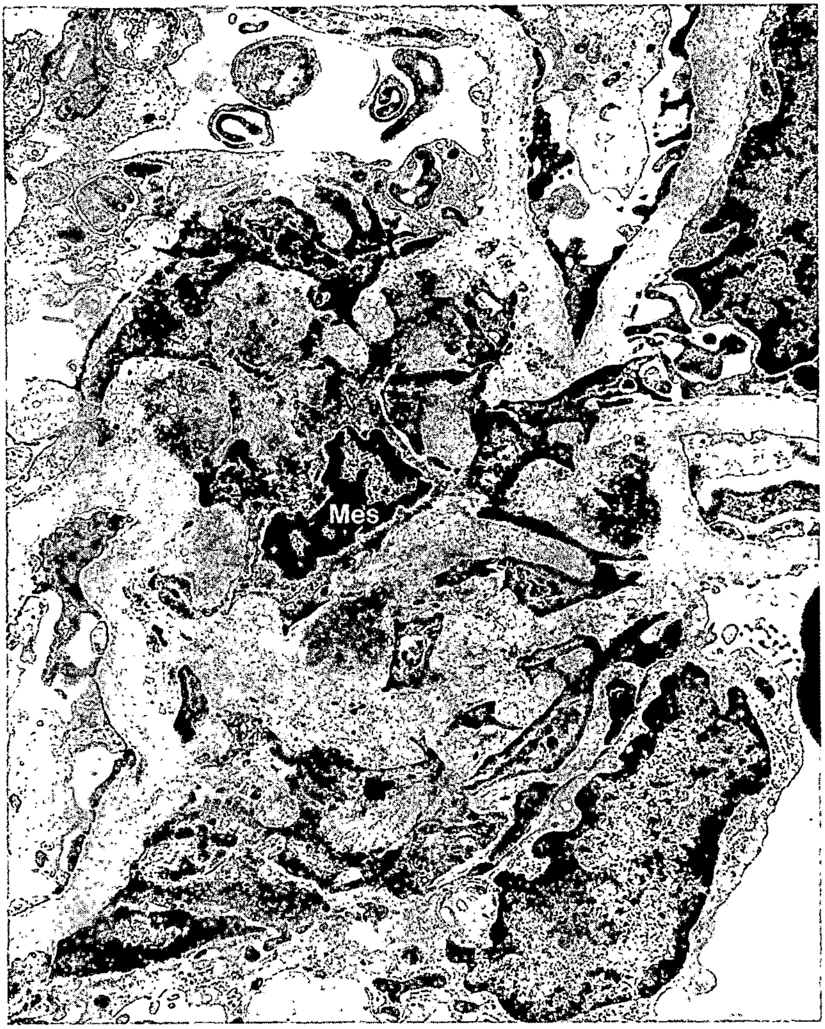


Figure 3a Dense deposits localized to the mesangium (Mes) seen by electron microscopy in Class II lupus nephritis.

findings usually signifying transformation to Class IV (see below). Creatinine clearance may be mildly diminished to 60–80 ml/min during periods of active disease. Small to moderate doses of corticosteroids (prednisone up to 0.5 mg/kg/day) generally reverse each of these clinical renal abnormalities. Often the dose of steroids employed has been that necessary to treat the accompanying extrarenal manifestations rather than the nephritis itself.

Some patients with Class III disease may have persistent proteinuria or hypocomplementemia despite therapy with escalating doses of steroids.

Clinical abnormalities seen in patients with DPLN are qualitatively similar to those with Class III, but are more profound. Presentation with or development of the nephrotic syndrome is common; proteinuria may exceed 10 g/24 hr. The urine sediment contains red and white blood cell casts as

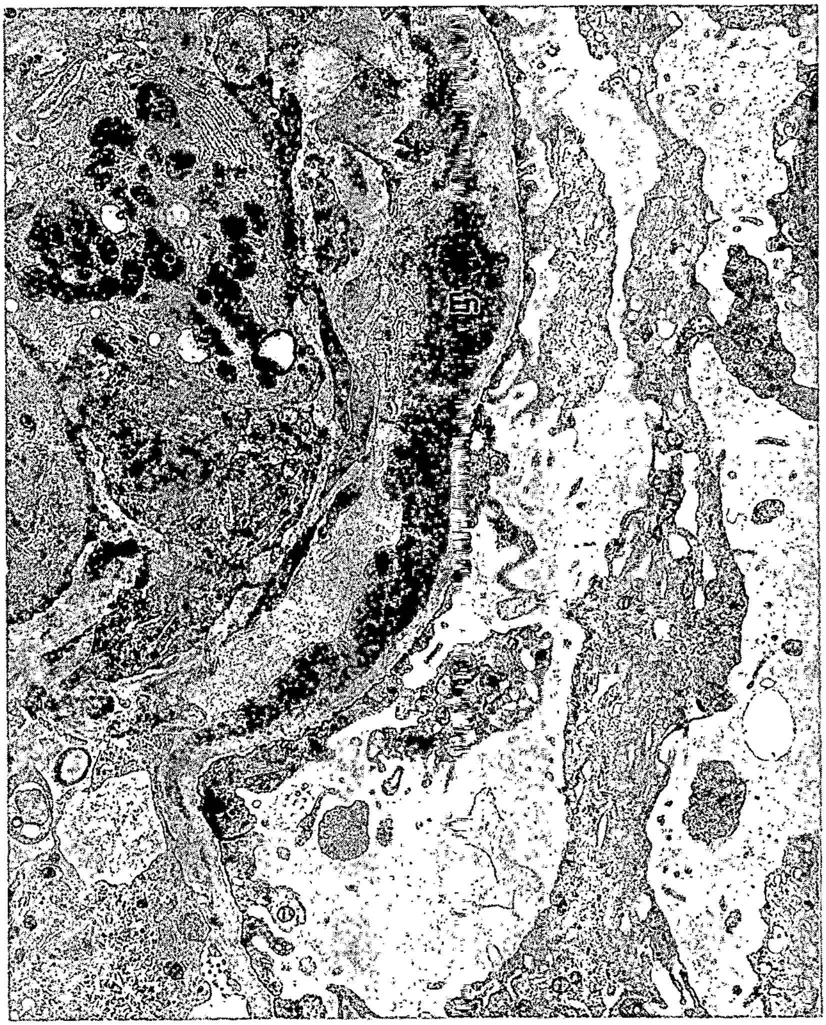


Figure 3b Dense deposits (En) present in a subendothelial location within the glomerular capillary basement membrane on electron microscopy. Subendothelial deposits are found in Class III and Class IV lupus nephritis.

well as many free red cells. More than half of Class IV patients have reduced creatinine clearance. Almost all studies of the natural history of the Class IV lesion are complicated by the introduction of steroid or other therapy. As emphasized by Baldwin et al, the untreated lesion will very likely progress to complete renal failure in less than two years after months of

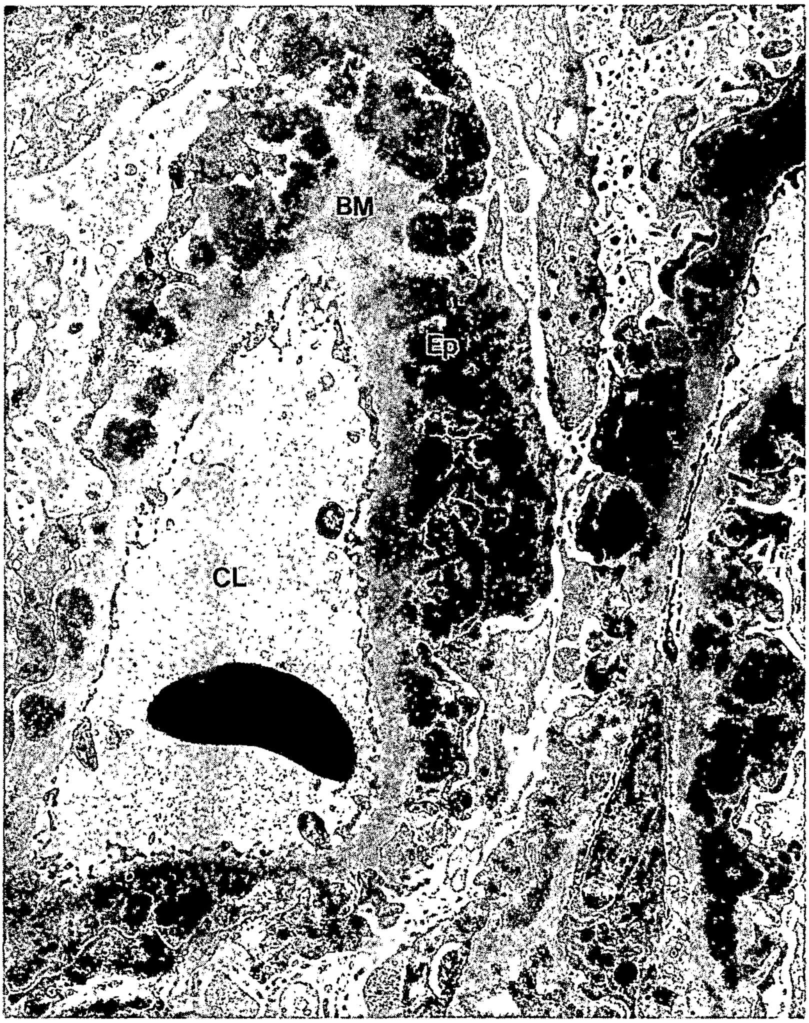


Figure 3c Electron dense deposits (Ep) present in a subepithelial location within the glomerular capillary basement membrane (BM) as seen by electron microscopy. CL represents the capillary lumen.

nephrotic syndrome, hematuria, and hypocomplementemia. Response to therapy in DPLN is variable and unpredictable, making the assessment of therapy difficult. Clinical abnormalities in Class IV may be reversed during the course of treatment, which results in long-term sustained remission. Some Class IV patients may have repeated exacerbations of active disease, each time responding to therapy, and ultimately maintaining normal renal function, although proteinuria and/or serologic abnormalities may persist. Others may progress despite intensive therapy to renal insufficiency requiring maintenance hemodialysis or renal transplantation.

Considerable attention has been focused upon the significance of persisting serologic abnormalities in patients with SLE, especially those with normal renal function. Lightfoot & Hughes (8), studying 16 patients with lupus nephritis, found that clinical remission lasted longer when serologic abnormalities were returned to normal by therapy, and that a more severe exacerbation of nephritis was likely to occur if anti-DNA titers persisted at a high level. Hypocomplementemia alone resolved spontaneously without a recognizable disease flare in nearly half the observed instances. Appel et al (9) prospectively studied the effect of normalizing the CH_{50} in 25 patients with lupus nephritis. Among those whose CH_{50} was maintained in the normal range for two years, urinary protein excretion decreased or remained low, and repeat renal biopsies showed improvement or stabilization of glomerular pathology. In patients whose CH_{50} remained abnormal despite tolerated doses of therapy, urinary protein excretion failed to decrease or worsened, and repeat renal biopsies in 5 of 6 patients showed worsening of the glomerular lesion. On the other hand, Gladman et al (10) report 14 patients with persistently low serum complement levels and high titers of anti-DNA antibodies, all of whom have remained asymptomatic without treatment for a mean period of more than 4 years. Clearly, the course of an individual patient with a Class IV lesion cannot be predicted with certainty.

Membranous lupus nephritis (Class V) appears to behave differently from the proliferative forms of disease. When assigning this classification to a given renal biopsy, Donadio et al (11) stress the necessity of excluding those lesions with proliferative features, even if the predominant change is basement membrane thickening. Likewise, it is essential to this classification that immune complexes be limited to a subepithelial or intramembranous location, and not to the subendothelium on EM. The nephrotic syndrome is the most prominent clinical feature in patients with Class V, present either initially or developing later in the course. Donadio et al found that progression to end-stage renal failure did not occur among 28 SLE patients over a period of up to 10 years, even in those patients whose nephrotic syndrome persisted, although one developed azotemia (creatinine 4.6

mg/dl). Baldwin et al (4), by contrast, observed a somewhat different course among 24 patients with a membranous lesion. When characterized by persistent nephrotic syndrome, a slow progression to renal failure ensued (6 of 17 patients) over a period of up to 6 years. This outcome was prevented in one third of patients in whom remission of the nephrotic syndrome could be achieved.

Eventual progression to renal failure in proliferative or membranous lupus nephritis is often characterized by deterioration attributable to the accumulation of salt, water, and nitrogenous wastes. Clinical problems in the lupus patient in renal failure are frequently related to severe hypertension, congestive heart failure, and fluid overload, as is true for uremia of other cause. Extrarenal manifestations of active SLE often disappear, and serologic abnormalities return toward normal, with an improvement in CH_{50} and C3 levels and decrease in the titer of antinuclear antibodies (12). The response to traditional therapy for lupus nephritis is generally ineffective and increases risk of iatrogenic infection. Uremia, while often immunosuppressive, may not always be totally effective in suppressing exacerbations of active lupus. We have observed three patients receiving maintenance hemodialysis for 1–4 years who continued to manifest exacerbations of severe extrarenal lupus including fever, arthritis, pericarditis, and gastrointestinal vasculitis, with increases in antinuclear antibody and anti-DNA titers, requiring moderate to large doses of steroid.

HISTOLOGIC TRANSFORMATIONS

When Baldwin et al originally proposed that lupus nephritis could be subdivided according to glomerular histologic findings, it was presumed that focal proliferative, diffuse proliferative, and membranous forms were separate and distinct entities that remain unchanged except for alterations in severity. The belief that the initial renal biopsy thereby predicted ultimate prognosis carried important therapeutic implications. In 1974, however, Ginzler et al (13) demonstrated that 9 of 31 patients with an initial renal biopsy showing mesangial or focal glomerulonephritis progressed to diffuse proliferative nephritis over a period of 1–12 years. Coincident with a transformation in histologic type, a deterioration in clinical renal status, represented by the appearance of new hypocomplementemia, azotemia, or significantly increased proteinuria was observed (13). Shortly thereafter, Zimmerman et al (14) reported a similar progression among 6 of 17 patients with an initial mesangial or focal glomerular lesion.

Prediction of the likelihood of progression at the time of initial biopsy is difficult. Ginzler and co-workers found no difference in clinical features at presentation or in the initial pattern or location of IF or EM deposits within the glomerulus to differentiate between the groups with and without progression. Zimmerman reported greater initial proteinuria and more dense subendothelial immune deposits among the patients who ultimately progressed to a more severe histologic lesion. Others have subsequently reported progression from mesangial or focal to diffuse nephritis in about 10–45% of patients with an initial mild lesion (4, 15, 16). In all cases, patients demonstrating progression to diffuse disease have followed a subsequent course compatible with the more severe lesion.

Transitions between the various proliferative lesions and membranous lupus nephritis have also been observed. In 1972 Hayslett et al (17) reported 7 patients with an initial "severe" proliferative glomerulonephritis who underwent marked decrease in intensity of the inflammatory reaction on LM and alteration in the histologic lesion with diffuse thickening of the capillary wall corresponding to epithelial and intramembranous deposits resembling membranous glomerulonephritis. These changes appeared to be in response to aggressive therapy with corticosteroids and azathioprine. Others have documented transformation from a focal or diffuse proliferative lesion to a membranous one (4, 15, 16).

Of perhaps greater significance is the observation that patients with true membranous (Class V) disease may undergo transformation to DPLN (Class IV). This change is generally heralded clinically by an abrupt increase in the degree of proteinuria and the appearance of azotemia (4, 15).

Finally, it is reassuring to note that further deterioration may not take place in all patients with DPLN; stabilization for years and even reversion to a mesangial lupus nephritis may occur. Baldwin et al (4) observed five such Class IV patients in sustained clinical remission who had repeat renal biopsies showing mesangial proliferation, with the development of some glomerulosclerosis. We are following a 34-year-old woman originally with severe but nonazotemic Class IV nephritis and intense subendothelial EM immune deposits who has been in clinical remission for five years, including reversion to a negative antinuclear antibody test, after treatment with azathioprine and prednisone. A second renal biopsy demonstrated complete hyalinization of about 15% of her glomeruli; however, the rest showed neither proliferative changes nor capillary basement membrane thickening. There was marked diminution in the quantity of EM deposits, with those present mostly in the mesangium and only a rare subendothelial deposit.

RESULTS OF THERAPEUTIC TRIALS

Following the introduction of adrenocorticosteroid therapy of SLE in the early 1950s, the literature suggests a consistent trend of improved survival corresponding to the use of escalating steroid doses (18–20). In some instances steroid therapy dramatically interdicted a rapidly downhill course. The use of corticosteroids by Pollack et al (21) in a regimen now considered to be low dose, generally 15–20 mg/day, did not alter survival in patients with lupus glomerulonephritis. When given in an uncontrolled trial of 60–100 mg/day for 6 months, however, greater survival and reduction in azotemia and improvement in renal histology was observed (21). In the most optimistic view of the efficacy of steroids, Lange et al (22) reported 100% survival during a mean of 21 months in 15 patients with lupus nephritis treated with sufficient prednisone (maximum 180 mg/day) to maintain a normal serum complement level.

There are many reports of the use of immunosuppressive and cytotoxic agents in the treatment of lupus nephritis, in the form of anecdotal cases, uncontrolled series, retrospective comparisons of two or more therapeutic regimens, and prospective controlled randomized trials. As employed in the suppression of human allograft rejection, azathioprine was the first such widely tried drug regimen. Though still prescribed, it has been incompletely evaluated, engendering debate regarding its therapeutic efficacy and toxicity. Early uncontrolled studies suggested that azathioprine in a dose of 1–3 mg/kg/day clinically improved azotemia and proteinuria in 50–100% of patients with severe lupus nephritis previously unresponsive to high doses of steroids (17, 23, 24). Sztejnbok et al (25) in 1971 reported a prospective randomized trial in which 16 patients receiving azathioprine (2.5 mg/kg) for 1–4 years had lessened mortality and morbidity, required less steroid, and maintained better renal function when compared to 19 patients treated with prednisone alone. This experience was confirmed and extended in a larger retrospective analysis at the same institution (26). Among 54 patients with Class IV or V nephritis and/or central nervous system lupus treated with azathioprine and followed for an average of 2.7 years, 15 (28%) died, whereas 10 of 14 (71%) similar patients who had never received azathioprine died during a mean follow-up of 1.3 years. Azathioprine therapy did not decrease mortality during the first six months after presentation with an acute exacerbation of severe disease, however. Seven of 23 (30%) azathioprine-treated patients died during this period, compared to 5 of 12 (41%) who did not receive this drug. Maintenance therapy with azathioprine in patients with normal kidneys or mild renal disease was associated with rare (1 of 20 patients) progression to Class IV or V nephritis. Patients who did not receive azathioprine or had discontinued it had a much higher incidence of progression (11 of 34). Recently, Barnett et al (27) observed that maintenance azathioprine treatment (1–4 mg/kg) in 47 patients with diffuse proliferative or membranous nephritis was associated with survival of 82% at 5 years and 74% at 10 years. That 3 of 4 patients with severe lupus nephritis may live a decade represents a substantial improvement over the 80%

5-year mortality observed by Baldwin et al in 1970 in patients treated with steroids alone (1).

Other investigators have expressed reservation about the therapeutic efficacy of azathioprine in lupus nephritis. Maher & Schreiner (28) failed to detect benefit of azathioprine in patients previously unresponsive to steroids, including those who were uremic when treatment was begun. The lack of value of maintenance azathioprine in addition to steroids was reported in a 2-year controlled trial by Hahn and coworkers (29), and in a similar 3-year study by Donadio et al (30). All of Donadio et al's 16 patients survived; in Hahn et al's series survival was similar among the 11 patients who received azathioprine and the 13 who did not.

Reluctance to prescribe azathioprine is based not only on its lack of proven efficacy but also on the many adverse reactions attributed to it, for example increased susceptibility to bacterial, viral, and fungal infections and an increased risk of malignancies, especially lymphomas, with long-term use (31–35). Most of the data on azathioprine-induced toxicity come from extensive experience in human renal transplant recipients (36–40). In contrast, Ginzler et al (41) failed to find an increased incidence of bacterial or fungal infection in a large series of azathioprine-treated patients followed prospectively for 345 patient-years. An early report of the appearance of cerebral lymphoma in a patient receiving azathioprine for SLE has not been corroborated (42).

The nitrogen mustard derivative cyclophosphamide has also been tested in several controlled trials in patients with lupus nephritis. A 10-week double-blind study at the NIH compared cyclophosphamide (average dose 1.17 mg/kg/day) to placebo in two groups of seven and six patients given up to 30 mg prednisone/per day (43). There was greater improvement in the cyclophosphamide-treated group as judged by anti-DNA antibody levels, serum complement, urine sediment, degree of proteinuria, and extrarenal manifestations of active lupus, but no difference in creatinine clearance was observed. Follow-up of the same patients for 28 months showed a gradual deterioration in renal function in both groups. One of the two deaths among cyclophosphamide-treated patients was due to opportunistic infection and the other to pulmonary embolism; 3 patients treated with prednisone had early nonrenal deaths and 2 eventually required maintenance hemodialysis for end-stage renal failure (44).

Donadio and co-workers (45) prospectively randomized 39 patients with DPLN to receive prednisone (average dose 40 mg/day), or prednisone (average 29 mg/day) and cyclophosphamide (average dose 107 mg/day), for 6 months. At the study's end there was no significant difference between the two groups with regard to the degree of improvement in creatinine clearance, proteinuria, or anti-DNA antibody titers, although the CH_{50}

levels were higher in the patients who had received cyclophosphamide. During a subsequent average follow-up of 24 months in the original patients and an additional 11 patients similarly randomized, however, the incidence of exacerbation in the cyclophosphamide group was lower (3 of 21) than in the control group (10 of 21) among patients who had achieved an initial disease remission (46).

Cyclophosphamide therapy can cause serious adverse reactions, especially when administered continuously for several months or longer. Toxicity in patients treated with this drug for nonmalignant disease includes alopecia, amenorrhea secondary to ovarian failure, lymphoproliferative disorders, and hemorrhagic cystitis with subsequent bladder fibrosis and carcinoma (47–50). Based on experiments in NZB/NZW mice in which "low" doses of azathioprine, cyclophosphamide, and prednisone (1 mg/kg each) acted synergistically to retard the development of naturally occurring nephritis (51, 52), a similar regimen was developed for human lupus nephritis. This attempt to minimize toxicity of azathioprine and cyclophosphamide by using them in combination, both in low dose (1.25 mg/kg), was disappointing. Ginzler et al (53) found no benefit of this regimen in addition to prednsione when compared to "full-dose" azathioprine (2.5 mg/kg) plus prednisone in a one-year double-blind trial in 14 patients with severe lupus nephritis. Furthermore, hemorrhagic cystitis and amenorrhea attributable to cyclophosphamide were not avoided with the lower dose of this drug.

From the trials of immunosuppressive drugs reported to date, no substantiated regimen can be proposed as either safe from side effects or of proven efficacy in arresting or reversing active lupus nephritis. Because of anecdotal accounts of dramatic reversals of lupus nephritis and NZB/NZW disease with cytotoxic or other drug regimens, and the sometimes desperate clinical circumstance of abrupt loss of renal function in young adults, the appeal of new drug programs continues. Currently both controlled and uncontrolled drug trials including new immunosuppressive drugs, modified treatment protocols for azathioprine in combination with agents such as heparin, heparin in combination with nonsteroidal antiinflammatory agents, and alternative measures for reducing immune responsiveness with or without drug therapy are under evaluation.

Kincaid-Smith (54) suggested that a mechanism for the obliteration of glomerular capillaries in a number of glomerulonephropathies, including lupus nephritis, involves subendothelial deposition of material containing not only immunoglobulins and complement but also fibrin, which can be demonstrated by immunofluorescent staining techniques. This has provided the theoretical basis for treating acute lupus nephritis with anticoagulants. Suc et al (55) reported a regimen of anticoagulation with heparin for 15–60 days, followed by indomethacin (100 mg/day) for its antiinflammatory and antiplatelet effects. Six patients with lupus nephritis received this regimen; three had total resolution of the nephrotic syndrome, and three had a partial remission, with proteinuria reduced to 1 mg/min or lower. Cade et al (56) also reported a regimen in which 13 patients with lupus nephritis received sufficient heparin subcutaneously to prolong the clotting time beyond 30 minutes, in combination with azathioprine (1–2 mg/kg). Twelve patients had a significant improvement in creatinine clearance after six months of therapy; however, one patient subsequently had an exacerbation of disease. Two renal deaths occurred in this group, and one patient had a severe retroperitoneal hemorrhage requiring heparin discontinuation.

Preliminary reports regarding the "immunoregulatory" agent Frentizole®, 1-(6-methoxy-2-benzothiazolyl)-3-phenyl urea, are equivocal. Although extrarenal manifestations of lupus such as rash, synovitis, mucosal ulcers, and pleurisy improved in four of six patients treated with the drug (57), neither creatinine clearance nor urinary protein excretion were altered favorably. In our experience with this agent in nine adults with SLE, low and probably suboptimal doses (1–3 mg/kg/day) caused considerable gastrointestinal toxicity, including epigastric pain, nausea, and reversible increases in alkaline phosphatase, often necessitating discontinuation of the drug before any clinical benefit was observed.

The complexities of the immune response include participation of both helper and suppressor thymus-derived lymphocytes. Stimulation of a suppressor lymphocyte might mute immunologically induced disease in SLE. Levamisole is classed as a stimulator of the immune response. Gordon & Yanigihara (58) treated 16 patients with levamisole and observed significant improvement in extrarenal disease manifestations. However the incidence of active nephritis was too small to assess the efficacy of the drug. Adverse reactions to levamisole were minimal, warranting a controlled prospective trial in lupus nephritis.

Intravenous "pulse" therapy with large doses of methylprednisolone (250–1000 mg/day for 3–12 days) has received considerable attention recently. Cathcart et al (59) inferred a rationale for pulse steroid therapy from the observation that histologic changes, including interstitial inflammation and edema, in the kidney of a lupus patient undergoing rapid renal deterioration resembled those in acute renal transplant rejection, an immune process purportedly reversed with large doses of intravenous steroids (59). Of seven patients with lupus nephritis and rapidly deteriorating renal function who were treated with a pulse steroid regimen, all had reversal of hypocomplementemia and decrease in anti-DNA titers; five had return to baseline serum creatinine levels within one month. Oral steroid therapy was continued at a level necessary to control extra-renal disease manifestations, in order to avoid the toxicity of continued high dose steroids. Dosa et al (60) similarly treated four patients, three with oliguric renal failure, and were enthusiastic about the results. Nevertheless, proteinuria was not significantly affected, creatinine clearance did not increase above 50 ml/min in those patients with previously impaired renal function, and subsequent exacerbations of active nephritis occurred in all four patients, necessitating repeat steroid pulses. The largest series of patients treated with intravenous pulse methylprednisolone was reported by Kimberly et al in 1979 (61). Of 28 patients, 25 with active renal disease, 17 had had recent deterioration in renal function. Four patients had a rapid, sustained improvement in renal manifestations, four developed end-stage renal failure requiring maintenance hemodialysis, and one died after eight weeks with continuing fulminant disease.

Our experience with pulse methylprednisolone in ten patients with progressive lupus nephritis has been similarly equivocal; no patient has had clear and sustained reversal of azotemic nephritis. Untoward effects of large doses of steroids include development of pulmonary edema presumably secondary to salt and water retention, and severe obtundation requiring emergency peritoneal dialysis in a patient whose blood urea nitrogen rose from about 60 to over 300 mg/dl within 24 hours after the pulse. As is true for the reversal of renal allograft rejection, the worth of bolus doses of intravenous methylprednisolone in lupus nephritis is by no means established.

Plasmapheresis, which in uncontrolled trials may improve or induce remission in patients with the glomerulonephritis of Goodpasture's syndrome (62, 63), has been advocated in the treatment of lupus nephritis. Modern plasmapheresis devices use a cell separator to remove serum rapidly, thus presumably extracting antigen-antibody complexes that may be responsible for inflammatory changes and subsequent organ damage. A major advantage of plasmapheresis may be the avoidance of toxic effects of immunosuppressive and antiinflammatory drugs, which may then be administered in lower doses. Verrier Jones and co-workers (64) used plasmapheresis in eight patients with SLE, four of whom had evidence of circulating immune complexes. A decrease in the level of immune complexes correlated partially with an increase in CH_{50} levels and improvement in clinical symptoms. The four patients with little evidence of circulating immune complexes failed to show convincing clinical improvement with plasmapheresis.

Other investigators are presently studying the efficacy of plasmapheresis, at least one in the form of a prospective controlled trial (W. Clark, personal communication), with promising early results. No long-term data are available to permit evaluation of sustained benefit. It must also be noted that the direct cost of plasmapheresis may approach $1000 per course, far exceeding that of any form of drug therapy. For the present, plasmapheresis must be regarded as a theoretically appealing, expensive, and unproven treatment for lupus nephritis.

THE NZB/NZW MOUSE AS A MODEL OF HUMAN LUPUS

The New Zealand Black (NZB) mouse, a strain developed for cancer research, was observed to develop spontaneously a group of "autoimmune" phenomena, including renal disease and autoimmune hemolytic anemia (65). Hybrids of the NZB mouse, and the New Zealand White (NZW) mouse result in an F1 generation with a modified syndrome in which hemolytic anemia is less apparant, but renal disease is universal. The disease is more severe in the female NZB/NZW mouse, with the majority dying with uremia between six and twelve months of age. Virtually all develop antinuclear antibodies, hematuria, and the nephrotic syndrome (66).

Light and electron microscopic examination of the NZB/NZW kidney reveals a progressive lesion beginning with subepithelial nodularity of the glomerular basement membrane and small dense deposits in the mesangium. By 24 weeks, there is fusion of epithelial foot processes around subepithelial nodules, progression of the deposits in the mesangium, subendothelial dense deposits, and further proliferation of capillary endothelial cells. The characteristic wire-loop lesion of human glomerulonephritis is rare in the mouse model (67).

The predictable development of immunologic and histologic abnormalities over a specific time course in the NZB/NZW mouse makes it a useful model for studying potentially beneficial therapeutic agents in human lupus nephritis. During the 1960s a number of investigators demonstrated that adrenocorticotropic hormone, corticosteroids including cortisone and betamethasone, and immunosuppressive agents including azathioprine and cyclophosphamide suppressed the renal lesion of NZB/NZW mice and prolonged survival (68). Subsequently, as previously mentioned, combinations of these agents were found to act synergistically in retarding development of the glomerulonephritis, while lower doses minimized the toxicity of each agent (51, 52).

A drawback to the NZB/NZW lesion as a model of human lupus nephritis is that patients with established lupus nephritis do not respond to immunosuppressive agents as predictably or as well as do mice. This must be remembered when evaluating other immunosuppressive agents screened in the murine model.

Newer immunoregulatory agents so far untested in controlled trials in humans with SLE have been used in murine disease. Thymosin, a fraction of thymus that improves some T-helper cell functions and restores normal thymocyte proliferative and suppressive function, failed to delay the appearance or severity of NZB/NZW nephritis (69). Levamisole delayed the development of antinuclear antibodies and proteinuria and extended survival (70), and when given after a course of cyclophosphamide delayed reappearance of antibodies to DNA and decreased glomerular immunoglobulin deposition (71).

Ribavirin, a broad spectrum antiviral agent, equaled cyclophosphamide in reducing anti-DNA antibodies, preventing proteinuria, and prolonging survival in NZB/NZW mice, when given in optimal dose (72). The effect of ribavirin is in keeping with the assumption that an environmental factor, presumably a virus, in combination with genetic susceptibility, is responsible for the development of the lupus-like syndrome in the NZB/NZW mouse. Viral antigen (Gross soluble antigen, GSA) occurs in NZB mice as they develop hemolytic anemia, then disappears when antibodies to GSA develop and renal disease appears (73). However, no such proof of viral replication in human lupus has been demonstrated. Microtubular structures resembling viral particles have been identified in the kidneys and other tissues of SLE patients; viruses from such tissues have not been definitely isolated (74–76).

Attempts have been made to alter the course of murine nephritis based on the knowledge that the disease is generally more severe in female mice, a similarity to the known predilection for females in the human disease. Steinberg et al (77) observed that castration of male NZB/NZW mice at five weeks slightly increased DNA binding at 6.5 months of age, and Roubinian et al (78) showed that castrated male mice developed a disease pattern similar to that of female mice, with premature mortality. They also demonstrated that castrated female mice given continuous androgen treatment had a decreased mortality compared to castrated females given replacement estrogen (14 vs 94% at 8 months). They failed to show an effect on life span of prepubertal castration of female NZB/NZW mice. Ducovic et al (79) demonstrated, however, that treatment of female NZB/NZW mice with the antiestrogen compound Nafoxidine® reduced anti-DNA antibodies, decreased proteinuria, and improved survival. At present hormonal manipulation does not appear practical in established human SLE, but an understanding of hormonal and immunologic interactions may help to clarify the pathogenesis of this disease.

THERAPEUTIC RECOMMENDATIONS

Several approaches to lupus nephritis have been advocated, but none of those so far applied in a controlled trial has been found to have clearcut advantages over another, including, perhaps, no therapy at all. Nearly all of the reports of clinical benefit in lupus nephritis are uncontrolled and anecdotal. Even those studies with prospective randomization of patients are generally too small to demonstrate a statistical difference between two therapeutic regimens. Furthermore, because SLE is a disease with such protean manifestations, the groups being compared were often not truly equatable. It has been suggested, in fact, that SLE may be a different disease in widely separated geographical areas, such as the US and Australia (80). These caveats imply that little conclusive knowledge has been gained from the large but admittedly disappointing previous experience. McShane et al (81) point out that the substantial literature on the treatment of lupus nephritis might be reanalyzed by applying "stratification techniques" both retrospectively and prospectively. The choice of therapy for the individual patient with SLE might be inferred not from the average outcome but from the outcome and best therapy for specifically similar patients. To achieve this end, many patients of differing disease severity and manifestations must be assigned to each treatment subset. Because such a broad patient spectrum cannot practically be achieved at one medical center, information science techniques, cooperative efforts, and access to large quantities of recorded prior experience are required.

Until such critically evaluated comparative data is routinely available, the present proposals for therapy in lupus nephritis must be viewed as a series of compromises between the physician's need to treat and the desire to do no harm. A balance between the need for a relatively simple uniform protocol that will yield further data and the necessity to weigh each patient's clinical and immunologic abnormalities and previous response to therapy must be achieved. With this in mind, the following recommendations are proposed:

1. In all forms of lupus nephritis, treatment of extrarenal manifestations of disease should not be ignored.
2. An attempt should be made initially to reverse the serologic and/or clinical abnormalities of urine sediment and protein excretion observed in patients with mesangial and focal glomerulonephritis. This can often be achieved with small to moderate doses of oral prednisone (up to 0.5 mg/kg). In patients who fail to respond after six to eight weeks or who develop clinically significant toxicity, steroids should be tapered over two to four weeks but not abruptly discontinued. Such patients should be observed carefully for signs of clinical deterioration that might suggest progression to a diffuse lesion, with its attendant poor prognosis.
3. Patients with a diffuse proliferative glomerular lesion and proteinuria and/or azotemia should initially be treated aggressively with high-dose oral steroids (1.0–1.5 mg/kg/day of prednisone). If no response or only partial

response occurs within four to six weeks, or if significant toxicity or clinical deterioration occurs sooner, additional therapy should be included, tapering steroids as indicated by toxicity. A predetermined, time-limited (twelve weeks) course of oral cyclophosphamide (2.0–2.5 mg/kg) may be instituted in patients whose serum creatinine is under 6 mg/dl and who have no evidence of chronic infection, but it should be discontinued as soon as clinical improvement is sustained, in order to minimize toxicity. Alternatively, patients with severe lupus nephritis may be enrolled in investigational protocols of new immunosuppressive drugs or techniques, providing that the treatment course has been defined and patient and staff understand their participation in an experiment. To justify the risk of a new therapy, patients must be carefully selected and strict adherence to the protocol must be met if usable data and maximal safety are to be achieved. As part of a controlled study, patients with a very recent onset of azotemia (less than one month) might be suitable for pulse methylprednisolone therapy, as might those with manifestations of severe interstitial and tubular disease. Plasmapheresis might be logically tried in patients with systemic vasculitis and proliferative glomerulonephritis with increasing azotemia, assumed to be secondary to immune complex deposition.

4. Patients with repeated exacerbations of nephritis or those whose clinical status deteriorates when steroids are tapered may have a trial of azathioprine (2.5 mg/kg). If leukopenia develops, the dose should be tapered so that the white blood count is maintained above $3000/\text{mm}^3$. While not generally useful in treating acute episodes of SLE, azathioprine appears to maintain remission in some patients, thereby minimizing future renal damage. Its steroid-sparing effect may serve to reduce rather than increase overall toxicity of a drug regimen.

5. In the severely ill, toxic, imminently moribund patient, use of an experimental drug or method in the absence of a standard protocol is justified. We would resort to plasmapheresis repeated thrice weekly in the very sick patient in whom rapid unexplained deterioration occurs.

6. Membranous glomerulonephritis with the nephrotic syndrome should be treated initially with a trial of prednisone, 1 mg/kg/day, for six weeks. Proteinuria leading to hypoproteinemia increases the risk of infection. Should proteinuria decrease to less than 2 g/day, sustained prednisone at a lower dose (0.5 mg/kg/day or 1.0 mg/kg on alternate days) should be continued. If proteinuria is not significantly diminished after six weeks, prednisone should be withdrawn and fluid overload managed with conservative measures such as diuretics and a salt-restricted diet.

7. A clinical team of nephrologist and rheumatologist should plan a strategy for each patient with lupus nephritis and azotemia. The team should define the limits of acceptable toxicity prior to beginning therapy,

so that treatment can be altered before it creates greater morbidity and mortality than would occur in untreated lupus nephritis. While most physicians managing lupus nephritis have seen an occasional patient with return of normal renal function and a sustained remission of severe azotemia following aggressive therapy, the likelihood of a good response in patients whose serum creatinine has reacher 3 mg/dl is small. An alternative to continued cytotoxic drug therapy should be considered in patients whose serum creatinine has risen to 4 mg/dl or higher. Facilities for maintenance hemodialysis are now ubiquitously available, allowing the conscious choice of withdrawal of toxic drug therapy in favor of progression to uremia, with excellent long-term survival (82). Many patients entered into a long-term dialysis program will undergo remission of their extrarenal manifestations of SLE, possibly as a result of partially reversed uremia. Patients who find maintenance hemodialysis to be an unacceptable life-style may be encouraged by the results of renal transplantation (83). But transplantation should not be attempted until the patient's condition has been stabilized and extrarenal manifestations of SLE have become quiescent.

ACKNOWLEDGMENT

We wish to thank Drs. Anthony D. Nicastri and Chun-Kuo Chen for providing the light microscopic, immunofluorescent, and electron microscopic illustrations of the patterns of lupus nephritis.

Literature Cited

1. Baldwin, D. S., Lowenstein, J., Rothfield, N. F., Gallo, G., McCluskey, R. T. 1970. The clinical course of the proliferative and membranous forms of lupus nephritis. *Ann. Intern. Med.* 73:929–42
2. McCluskey, R. T. 1975. Lupus nephritis. In *Kidney Pathology Decennial 1966–1975,* ed. S. C. Sommers, New York: Appleton-Century-Crofts. 435 pp.
3. Brentjens, J. R., Sepulveda, M., Baliah, T., Bentzel, C., Erlanger, B. F., Elwood, C., Montes, M., Hsu, K. C., Andres, G. A. 1975. Interstitial immune complex nephritis in patients with systemic lupus erythematosus. *Kidney Int.* 7:342–50
4. Baldwin, D. S., Gluck, M. C., Lowenstein, J., Gallo, G. R. 1977. Lupus nephritis. Clinical course as related to morphologic forms and their transitions. *Am. J. Med.* 62:12–30
5. Bennett, W. M., Bardana, E. J., Houghton, D. C., Pirofsky, B., Striker, G. D. 1977. Silent renal involvement in systemic lupus erythematosus. *Int. Arch. Allergy Appl. Immunol.* 55:420–28
6. Mahajan, S. K., Ordonez, N. G., Feitelson, P. J., Lim, V. S., Spargo, B. H., Katz, A. I. 1977. Lupus nephropathy without clinical renal involvement. *Medicine* 56:493–501
7. Zweiman, B., Kornblum, J., Cornog, J., Hildreth, E. A. 1968. The prognosis of lupus nephritis: Role of clinical-pathologic correlations. *Ann. Intern. Med.* 69:441–62
8. Lightfoot, R. W. Jr., Hughes, G. R. V. 1976. Significance of persisting serologic abnormalities in SLE. *Arthritis Rheum.* 19:837–43
9. Appel, A. E., Sablay, L. B., Golden, R. L., Barland, P., Grayzel, A. I., Bank, N. 1978. The effect of normalization of serum complement and anti-DNA antibody on the course of lupus nephritis. *Am. J. Med.* 64:274–83
10. Gladman, D. D., Urowitz, M. B., Keystone, E. C. 1979. Serologically active clinically quiescent systemic lupus erythematosus. A discordance between

clinical and serologic features. *Am. J. Med.* 66:210–15

11. Donadio, J. V. Jr., Burgess, J. H., Holley, K. E. 1977. Membranous lupus nephropathy: A clinicopathologic study. *Medicine* 56:527–36
12. Fries, J. F., Powers, R., Kempson, R. L. 1974. Late-stage lupus nephropathy. *J. Rheum.* 1:166–175
13. Ginzler, E. M., Nicastri, A. D., Chen, C. K., Friedman, E. A., Diamond, H. S., Kaplan, D. 1974. Progression of mesangial and focal to diffuse lupus nephritis. *N. Engl. J. Med.* 291:693–96
14. Zimmerman, S. W., Jenkins, P. G., Shelp, W. D., Bloodworth, J. M. B. Jr., Burkholder, P. M. 1975. Progression for minimal or focal to diffuse proliferative lupus nephritis. *Lab. Invest.* 32:665–73
15. Appel, G. B., Silva, F. G., Pirani, C. L., Meltzer, J. I., Estes, D. 1978. Renal involvement in systemic lupus erythematosus (SLE): A study of 56 patients emphasizing histologic classification. *Medicine* 57:371–410
16. Mahajan, S. K., Ordonez, N. G., Spargo, B. H., Katz, A. I. 1978. Changing histology patterns in lupus nephropathy. *Clin. Neph.* 10:1–8
17. Hayslett, J. P., Kashgarian, M., Cook, C. D., Spargo, B. H. 1972. The effect of azathioprine on lupus glomerulonephritis. *Medicine* 51:393–412
18. Dubois, E. L., Commons, R. R., Starr, P., Stein, C. S. Jr., Morrison, R. 1952. Corticotropin and cortisone treatment for systemic lupus erythematosus. *J. Am. Med. Assoc.* 149:995–1002
19. Merrell, M., Shulman, L. E. 1955. Determination of prognosis in chronic disease, illustrated by systemic lupus erythematosus. *J. Chron. Dis.* 1:12–32
20. Medical Research Council. 1961. Treatment of systemic lupus erythematosus with steroids. *Br. Med. J.* 2:915–18
21. Pollack, V. E., Pirani, C. L., Schwartz, F. D. 1964. The natural history of the renal manifestations of systemic lupus erythematosus. *J. Lab. Clin. Med.* 63: 537–50
22. Lange, K., Ores, R., Strauss, W., Wachstein, M. 1965. Steroid therapy of systemic lupus erythematosus based on immunologic considerations. *Arthritis Rheum.* 8:244–59
23. Drinkard, J. P., Stanley, T. M., Dornfeld, L., Austin, R. C., Barnett, E. V., Pearson, C. M., Vernier, R. L., Adams, D. A., Latta, H., Gonick, H. C. 1970. Azathioprine and prednisone in the treatment of adults with lupus nephritis. *Medicine* 49:411–32
24. Shelp, W. D., Bloodworth, J. M. B. Jr., Rieselbach, R. E. 1971. Effect of azathioprine on renal histology and function in lupus nephritis. *Arch. Intern. Med.* 128:566–73
25. Sztejnbok, M., Stewart, A., Diamond, H., Kaplan, D. 1971. Azathioprine in the treatment of systemic lupus erythematosus. A controlled study. *Arthritis Rheum.* 14:639–45
26. Ginzler, E., Sharon, E., Diamond, H., Kaplan, D. 1975. Long-term maintenance therapy with azathioprine in systemic lupus erythematosus. *Arthritis Rheum.* 18:27–35
27. Barnett, E. V., Dornfeld, L., Lee, D. B. N., Liebling, M. R. 1978. Long-term survival of lupus nephritis patients treated with azathioprine and prednisone. *J. Rheum.* 5:275–87
28. Maher, J. F., Schreiner, G. E. 1970. Treatment of lupus nephritis with azathioprine. *Arch. Intern. Med.* 125: 293–98
29. Hahn, B., Kantor, O., Osterland, K. 1975. Azathioprine plus prednisone compared with prednisone alone in the treatment of SLE. *Ann. Intern. Med.* 83:592–605
30. Donadio, J. V., Holley, K. E., Wagoner, R. D., Fergusen, R. H., McDuffie, F. C. 1974. Further observations in the treatment of lupus nephritis with prednisone and combined prednisone and azathioprine. *Arthritis Rheum.* 17:573–81
31. Steinberg, A. D., Plotz, P. H., Wolf, S. M., Wong, V. G., Agus, S. G., Decker, J. L. 1972. Cytotoxic drugs in treatment of nonmalignant diseases. *Ann. Intern. Med.* 76:619–42
32. Winner, H. I., Bagshawe, K. D., Coleman, J. C., Gerken, A. 1970. Microbiologic aspects of cytotoxic drug therapy. *Proc. R. Soc. Med.* 63:1068–70
33. Robson, G. S., MacKay, I. R. 1969. Generalized cytomegalovirus infection in a patient with lupoid hepatitis. *Australas. Ann. Med.* 18:147–50
34. Nogueira, J. R., Freedman, M. A. 1972. Acute pancreatitis as a complication of Imuran therapy in regional enteritis. *Gastroenterology* 62:1040–41
35. Sparberg, M., Simon, N., Del Greco, F. 1969. Intrahepatic cholestasis due to azathioprine. *Gastroenterology* 57: 439–41
36. Chandler, J. W., Kalina, R. E., Milam, D. F. 1972. Coccidioidal choroiditis following renal transplantation. *Am. J. Ophthalmol.* 74:1080–85

37. Schneck, S. A., Penn, I. 1971. De-novo brain tumours in renal transplant recipients. *Lancet* 1:983–86
38. Doak, P. B., Montgomerie, J. Z., North, J. D. K., Smith, F. 1968. *Br. Med. J.* 4:746–48
39. Tallent, M. B., Simmons, R. L., Najarian, J. S. 1970. Birth defects in child of male recipient of kidney transplant. *J. Am. Med. Assoc.* 211:1854–55
40. Malkzedah, M. H., Grushkin, C. M., Wright, H. T. Jr., Fine, R. N. 1972. Hepatic dysfunction after renal transplantation in children. *J. Pediatr.* 81: 279–85
41. Ginzler, E., Diamond, H., Kaplan, D., Weiner, M., Schlesinger, M., Seleznick, M. 1978. Computer analysis of factors influencing frequency of infection in systemic lupus erythematosus. *Arthritis Rheum.* 21:37–44
42. Lipsmeyer, E. A. 1972. Development of malignant cerebral lymphoma in a patient with systemic lupus erythematosus treated with immunosuppression. *Arthritis Rheum.* 15:183–86
43. Steinberg, A. D., Kaltreider, H. B., Staples, P. J., Goetzel, E. J., Talal, N., Decker, J. L. 1971. Cyclophosphamide in lupus nephritis: A controlled trial. *Ann. Intern. Med.* 75:165–71
44. Decker, J. L., Klippel, J. H., Plotz, P. H., Steinberg, A. D. 1975. Cyclophosphamide or azathioprine in lupus nephritis. A controlled trial: results at 28 months. *Ann. Intern. Med.* 83:606–16
45. Donadio, J. V. Jr., Holley, K. E., Ferguson, R. H., Ilstrup, D. M. 1976. Progressive lupus glomerulonephritis. Treatment with prednisone and combined prednisone and cyclophosphamide. *Mayo Clin. Proc.* 51:484–94
46. Donadio, J. V. Jr., Holley, K. E., Ferguson, R. H., Ilstrup, D. M. 1978. Treatment of diffuse proliferative lupus nephritis with prednisone and combined prednisone and cyclophosphamide. *N. Engl. J. Med.* 299:1151–55
47. Aptekar, R. G., Atkinson, J. P., Decker, J. L., Wolff, S. M., Chu, E. W. 1973. Bladder toxicity with chronic oral cyclophosphamide therapy in nonmalignant disease. *Arthritis Rheum.* 16:461–67
48. Warne, G. L., Fairley, K. F., Hobbs, J. B., Martin, F. I. R. 1973. Cyclophosphamide-induced ovarian failure. *N. Engl. J. Med.* 289:1159–62
49. Tannenbaum, H., Schur, P. H. 1974. Development of reticulum cell sarcoma during cyclophosphamide therapy. *Arthritis Rheum.* 17:15–18
50. Pollock, B. H., Barr, J. H. Jr., Stolzer, B. L., Eisenbeis, C. H. Jr., Agarwal, A., Margolis, H. M. 1973. Neoplasia and cyclophosphamide. *Arthritis Rheum.* 16:524–26
51. Gelfand, M. C., Steinberg, A. D., Nagle, R., Knepshield, J. H. 1972. Therapeutic studies in NZB/W mice. I. Synergy of azathioprine, cyclophosphamide, and methylprednisolone in combination. *Arthritis Rheum.* 15:239–46
52. Hahn, B. H., Bagby, M. K., Hamilton, T. R., Osterland, C. K. 1973. Comparison of therapeutic and immunosuppressive effects of azathioprine, methylprednisolone, and combined therapy in NZB/NZW mice. *Arthritis Rheum.* 6:163–70
53. Ginzler, E., Diamond, H., Guttadauria, M., Kaplan, D. 1976. Prednisone and azathioprine compared to prednisone plus low-dose azathioprine and cyclophosphamide in the treatment of lupus nephritis. *Arthritis Rheum.* 19:693–99
54. Kincaid-Smith, P. 1973. The role of coagulation in the obliteration of glomerular capillaries. In *Glomerulonephritis,* ed. P. Kincaid-Smith, T. H. Mathew, E. Lovell Becker, pp. 871–90 New York: Wiley
55. Suc, J. M., Conte, J., Mignon-Conte, M. 1973. Treatment of glomerulonephritis with indomethacin and heparin. See Ref. 54 pp. 927–47
56. Cade, R., Spooner, G., Schlein, E., Pickering, M., DeQuesada, A., Holcomb, A., Juncos, L., Richard, G., Shires, D., Levin, D., Hackett, R., Free, I, Hunt, R., Fregly, M. 1973. Comparison of azathioprine, prednisone, and heparin alone or combined in treating lupus nephritis. *Nephron* 10:37–56
57. Valentine, T. V., Kay, D. R. Walker, S. E., Bole, G. B. 1978. Frentizole therapy of active systemic lupus erythematosus. *Arthritis Rheum.* 21:599 (Abstr.)
58. Gordon, B. E. II, Yanagihara, R. 1977. Treatment of systemic lupus erythematosus with the T-cell immunopotentiator levamisole: A follow-up report of 16 patients under treatment for a minimum period of four months. *Ann. Allergy* 39:277–36
59. Cathcart, E. S., Scheinberg, M. A., Idelson, S. A., Couser, W. G. 1976. Beneficial effects of methylprednisolone "pulse" therapy in diffuse proliferative lupus nephritis. *Lancet* 1:163–66
60. Dosa, S., Cairns, S. A., Lawler, W., Mallick, N. P., Slotki, I. N. 1978. The treatment of lupus nephritis by methyl

prednisolone pulse therapy. *Postgrad. Med. J.* 54:628–32

61. Kimberly, R., Lockshin, M., Sherman, R., McDougal, J. S., Inman, R., Christian, C. L. 1979. Clinical efficacy of high-dose intravenous methylprednisolone pulse therapy. *Arthritis Rheum.* 22:629 (Abstr.)
62. Johnson, J. P., Whitman, W., Briggs, W. A. 1978. Plasmapheresis and immunosuppressive agents in antibasement membrane antibody-induced Goodpasture's syndrome. *Am. J. Med.* 64:354–59
63. Cove-Smith, J. R., McLeod, A. A., Blamly, R. W., Knapp, M. S., Reeves, W. G., Wilson, C. B. 1978. Transplantation, immunosuppression, and plasmapheresis in Goodpasture's syndrome. *Clin. Nephrol.* 9:126–28
64. Verrier Jones, J., Fraser, I. D., Bothamley, J., Davis, P., Hamblin, T. J. 1979. A therapeutic role for plasmapheresis in the management of acute systemic lupus erythematosus. *Plasma Ther.* 1:33–41
65. Bielschowsky, M., Helyer, B. J., Howie, J. B. 1959. Spontaneous hemolytic anemia in mice of the NZB/BL strain. *Proc. Univ. Otage Med. Sch.* 37:9–11
66. Helyer, B. J., Howie, J. B. 1963. Renal disease associated with positive lupus erythematosus tests in a cross bred strain of mice. *Nature* 197:197
67. Comerford, F. R., Cohen, A. S., Desai, R. G. 1968. The evolution of the glomerular lesion in NZB mice. A light and electron microscopic study. *Lab. Invest.* 19:643–51
68. Howie, J. B., Helyer, B. J. 1968. The immunology and pathology of NZB mice. *Adv. Immunol.* 9:215–66
69. Gershwin, M. E., Steinberg, A. D., Ahmed, A., Derkay, C. 1976. Study of thymic factors. II. Failure of thymosin to alter the natural history of NZB and NZB/NZW mice. *Arthritis Rheum.* 19:862–66
70. Russell, A. S. 1976. Therapeutic trials with levamisole and other agents in NZB/W mice. *J. Rheum.* 3:380–83
71. Zulman, J., Michalski, J., McCombs, C., Greenspan, J., Talal, N. 1978. Levamisole maintains cyclophosphamide-induced remission in murine lupus erythematosus. *Clin. Exp. Immunol.* 31:321–27
72. Klassen, L. W., Williams, G. W., Reinersten, J. L., Gerber, N. L., Steinberg, A. D. 1979. Ribavirin treatment in murine autoimmune disease. I. Therapeutic efficacy and effect on immune response. *Arthritis Rheum.* 22:145–54
73. Mellors, R. C., Aoki, T., Huebner, R. J. 1968. Further implication of murine leukemia-like virus in the disorders of NZB mice. *J. Exp. Med.* 129:1045–62
74. Fresco, R. 1970. Virus-like particles in systemic lupus erythematosus (Letter). *N. Engl. J. Med.* 283:1231
75. Gyorkey, F., Sinkovics, J. G., Min, K. W., Gyorkey, P. 1972. A morphologic study on the occurrence and distribution of structures resembling viral nucleocapsids in collagen diseases. *Am. J. Med.* 53:148–58
76. Tisher, C. G., Kelso, H. B., Robinson, R. R., Gunnells, J. C., Burkholder, P. M. 1971. Intraendothelial inclusions in kidneys of patients with SLE. *Ann. Intern. Med.* 75:537–47
77. Steinberg, A. D., Pincus, T., Talal, N. 1971. The pathogenesis of autoimmunity in New Zealand mice. III. Factors influencing the formation of antinucleic acid antibodies. *Immunology* 20:523–31
78. Roubinian, J. R., Papoian, R., Talal, N. 1977. Androgenic hormones modulate autoantibody responses and improve survival in murine lupus. *J. Clin. Invest.* 59:1066–70
79. Ducovic, M., Steinberg, A. D., Klassen, L. W. 1978. Effect of the antiestrogen, Nafoxidine, on NZB/W autoimmune disease. *Arthritis Rheum.* 21:414–17
80. Kincaid-Smith, P. 1973. Lupus nephritis. See Ref. 54, p. 1223
81. McShane, D. J., Porta, J., Fries, J. F. 1978. Comparison of therapy in severe systemic lupus erythematosus employing stratification techniques. *J. Rheum.* 5:51–58
82. Maxey, R. W., Rao, T. K. S., Butt, K. M. H., Ginzler, E., Kountz, S. L., Friedman, E. A. 1978. Hemodialysis and transplantation in patients with uremia due to systemic lupus erythematosus. *Artif. Organs* 2S:470–72
83. Amend, W., Vincenti, F., Covey, C., Epstein, W., Feduska, N., Salvatierra, O. Jr. 1977. Renal transplantation in systemic lupus erythematosus. *Proc. Dialysis Transplant Forum* 7:18–21

Ann. Rev. Med. 1980. 31:489–508
Copyright © 1980 by Annual Reviews Inc. All rights reserved

CONTROL OF CORONARY BLOOD FLOW

❖7375

Francis J. Klocke, M.D. and Avery K. Ellis, M.D., Ph.D.

Department of Medicine, State University of New York at Buffalo,
and Erie County Medical Center, Buffalo, New York 14215

FACTORS AFFECTING THE MYOCARDIAL O_2 DEMAND-SUPPLY RELATIONSHIP

The primary function of the coronary circulation is to supply the myocardium with its oxygen (O_2) requirements. Since the heart has only a limited and short-lived capacity for anerobic metabolism, the metabolic needs of the myocardium can be considered almost solely in terms of oxidative metabolism. Therefore, there is a very close relationship between O_2 demand and O_2 supply. The three major factors affecting demand are wall stress, contractility, and heart rate (1). Wall stress is directly proportional to blood pressure and radius of curvature and inversely proportional to wall thickness. Systolic blood pressure, either peak or mean, is often used as an estimate for myocardial wall stress, and increases in systolic pressure are associated with similar increases in wall stress. However, changes in either radius of curvature or wall thickness can bring about changes in wall stress independently of changes in pressure. These latter factors are potentially important, especially when comparing different individuals or when evaluating wall stress in a single patient before and after an intervention. In spite of these limitations, systolic pressure remains a useful estimate of myocardial wall stress. Wall tension is less frequently substituted for wall stress. While the calculation of tension does take chamber radius into account, it pertains only to thin-wall structures and neglects effects of wall thickness.

Contractility, the second major determinant of myocardial oxygen demand, is difficult to assess directly or quantitatively in man. In the intact heart sympathetic stimulation, catecholamine, or calcium administration can result in a substantial increase in myocardial O_2 uptake ($M\dot{V}O_2$) related

0066-4219/80/0401-0489$01.00

solely to the increased velocity of contraction (2–4). Conventional hemodynamic indices of contractility have important limitations and are ordinarily ignored in clinical estimates of $M\dot{V}O_2$ from hemodynamic parameters. Heart rate, the third important determinant of myocardial O_2 demand, is itself associated with inotropic effects. However, it is not clear whether the effect of rate on $M\dot{V}O_2$ is due solely to an increase in the number of contractions per minute, or to the concomitant increase in velocity of contraction as well (1). Other factors affecting $M\dot{V}O_2$, such as increases in stroke volume ("volume work"), are relatively minor, e.g. a 50% increase in stroke volume increases $M\dot{V}O_2$ by only ~4% (1).

Attempts in man to correlate specific hemodynamic parameters reflecting myocardial O_2 demand with direct measurements of O_2 consumption have been limited largely by the invasive nature of techniques for quantitating coronary blood flow. Studies confined to basal conditions in the cardiac catheterization laboratory have covered a limited range of O_2 demand. Direct measurements of myocardial O_2 consumption in man during exercise have been reported by groups from Minnesota (5–7) and Göteborg (8) using direct quantitation of coronary blood flow and the coronary arterial-venous O_2 difference. The Minnesota group studied normal young males at 2–3 levels of exertion using upright bicycle ergometry. The highest levels of exercise achieved corresponded to 85% of maximal heart rate and 56% of maximal total body oxygen consumption. The arterial-venous O_2 difference across the coronary bed increased by 10–15%, and $M\dot{V}O_2$ by 300–400% (assuming usual resting values of this parameter). Indices of myocardial O_2 consumption involving the product of systolic pressure and heart rate correlated with coronary blood flow ($r = 0.87–0.89$) and O_2 consumption ($r = 0.90$) better than tension-time index, heart rate, etc. The usefulness of "double product" as an index of myocardial oxygen consumption in groups of healthy, normal individuals was therefore established. However, 95% confidence limits for estimates of $M\dot{V}O_2$ from individual values of double product were substantial, i.e. ±8 ml O_2/100 g LV/min (Figure 3 in Reference 7).

The Göteborg group (8) studied patients who were supine rather than upright, again using a bicycle ergometer. The double product (DP) of systolic pressure and heart rate was in the same range as in the Minnesota studies and was again reported to be a useful index of myocardial O_2 uptake in control subjects. However, the slope of the regression line calculated for these subjects (See Reference 8, Table 2, patients 11–21) is less than the slope of the similar line for the entire Minnesota group of normal individuals ($M\dot{V}O_2 = 0.097$ DP − 0.62, $r = 0.89$ vs $M\dot{V}O_2 = 0.16$ DP − 3.9, $r = 0.85$). Whether the difference in slope can be explained by the supine vs upright position is unclear. The 95% confidence limits for prediction of

O_2 uptake from individual measurements of double product again vary widely. Similar data reported by the Göteborg group for patients with coronary artery disease seem difficult to interpret in light of potential limitations in this setting of the Xe^{133} technique used to measure coronary flow (9).

The Minnesota group repeated studies in normal male volunteers before and after beta blockade, again using upright bicycle exercise (7). Their findings emphasize that relative metabolic loads for the whole body and heart are determined separately, and do not necessarily change in parallel with a given intervention. Following systemic beta blockade with propranolol, a significantly higher workload was required to achieve a heart rate of 120 beats/min, although double product and myocardial O_2 consumption were the same at the two workloads. Similarly, when exercise to a given workload was repeated after propranolol, the workload was achieved at significantly lower values of double product and $M\dot{V}O_2$. The rate-pressure product continued to correlate well with myocardial O_2 consumption, despite the reduction in contractility produced by this agent. The continued correlation may, at least in part, have reflected "cancelling" effects of directionally opposite influences on O_2 consumption, e.g. reduced contractility secondary to propranolol vs increased wall stress secondary to a larger heart size during exercise. Also noteworthy was an 18% increase in average coronary arterial-venous O_2 extraction at the same level of double product (16.1 ml O_2/100 ml with propranolol vs 13.8 ml O_2/100 ml prior to propranolol).

Physical training can also affect myocardial O_2 demand. Myocardial O_2 requirements are clearly less for any given level of exercise in the trained individual. The diminished heart rate during exercise is an important factor in this response, while effects of hypertrophy and/or cavity dilation (10) are more difficult to sort out.

DRIVING PRESSURE AND RESISTANCE IN THE CORONARY VASCULAR BED

Oxygen supply to the myocardium depends on coronary blood flow and the coronary arterial-venous O_2 difference. About 65–85% of the O_2 contained in arterial blood is normally extracted by the myocardium in a single passage through the coronary bed. Accordingly, it is difficult for the heart to adjust to increased metabolic demand by increasing O_2 extraction (although, as noted above, a modest degree of increased extraction can occur during intense exercise or beta blockade). Changes in coronary blood flow are the primary mechanism by which the coronary circulation adjusts to changing myocardial O_2 demand. To understand the factors regulating coronary flow, one must consider in some detail the relationships among coronary flow, driving pressure, and vascular resistance.

The relationship between flow ($\dot{Q}$), pressure, and resistance in the left ventricular bed has most frequently been expressed using the traditional Ohm's Law formulation, $\dot{Q} = \Delta P/R$, where ΔP = driving pressure and R = vascular resistance. Resistance is potentially multifactorial and is discussed in depth below. The driving pressure ΔP is more complex than previously appreciated, and approaches for employing Ohm's Law need to be reexamined. Figure 1 schematically represents traditional concepts of pressure-flow relations within the coronary vascular bed. The solid black line depicts the relationship between flow and pressure over a wide range of these values. The range of pressure over which flow remains relatively constant—the "autoregulatory" range (11)—lies between the bounds of maximum vasodilation on the left and maximum vasoconstriction on the right. The black dot represents a typical operating point within this range. Vascular resistance calculated for the point usually employs the Ohm's Law formulation and places the point on one of the three dotted pressure-flow lines, depending on the value chosen for ΔP. Most often ΔP has been taken as the difference between inflow and right atrial pressures, or as inflow pressure alone. Some investigators have used mean pressure for the entire cardiac cycle while others have chosen mean diastolic values because of the

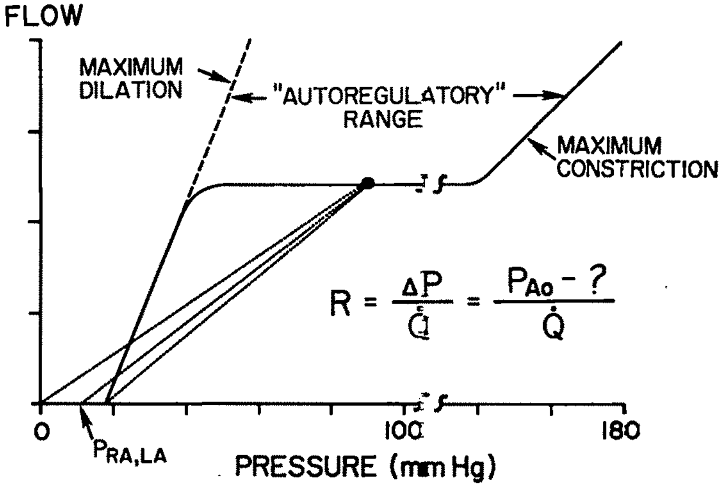


Figure 1 Pressure-flow relationships in the coronary bed. Dotted lines illustrate different calculations of resistance, depending on the choice of "back pressure" (see text for details): R = resistance, ΔP = driving pressure, $\dot{Q}$ = flow, P_{AO} = aortic (inflow) pressure, $P_{RA,LA}$ = right and left atrial pressures. (Reprinted with permission from *Anesthesiology*.)

preponderance of coronary flow during diastole. During the past few years attention has focused increasingly on the difference between inflow and left ventricular (or left atrial) diastolic pressures since local tissue pressure may limit flow even during diastole (12). Previous studies of autoregulation (e.g. 11) suggest that the minimum pressure required for any flow during maximum vasodilation (the intersection of the solid pressure-flow curve with the pressure axis) is actually higher than either right or left atrial pressures. Thus, the proper choice of driving pressure is at issue. Overestimates of ΔP will, of course, lead to overestimates of coronary vascular resistance.

Recently obtained data, schematically represented in Figure 2, indicate that the magnitude of this problem is greater than previously suspected. When instantaneous pressure-flow relationships are quantified during single long diastoles, flow consistently appears to be a linear function of inflow pressure (13, 13a). The reciprocal of the pressure-flow line provides a direct measure of coronary vascular resistance, and the pressure-axis intercept an estimate of the minimum inflow pressure required for any diastolic perfusion. Bellamy (13) reported the latter, here termed zero-flow pressure (P_{ZF}), to be ~40 mm Hg in conscious dogs under basal conditions; our laboratory recently verified it to be 3–5 times greater than mean left or right atrial pressure in open-chest dogs (13a). Thus, forces opposing diastolic flow

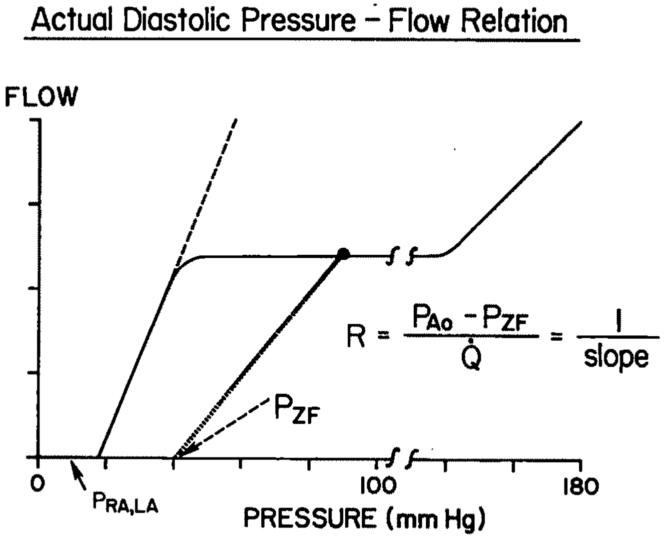


Figure 2 Pressure-flow relationship during a single diastole. The thick dotted line illustrates that the relationship is linear, with a pressure-axis intercept that is higher than any of the values for "back pressure" suggested in Figure 1: P_{ZF} = zero-flow pressure; other abbreviations as in Figure 1. (Reprinted with permission from *Anesthesiology*.)

in the immediate vicinity of intramyocardial vascular channels are appreciably higher than previously considered, and coronary vascular resistance can be significantly overestimated if P_{ZF} is ignored. In addition, this quantitatively important "back pressure" to flow can be changed by physiological interventions and probably plays a dynamic rather than static role in coronary flow regulation. These points are discussed further in the section on coronary reserve. The basic mechanism(s) by which P_{ZF} is controlled remains to be defined but is of considerable interest.

FUNCTIONAL COMPONENTS AND CONTROL OF CORONARY VASCULAR RESISTANCE

Functional Components

Coronary vascular resistance can be modeled as the sum of three functional components (14). (*a*) So-called basal viscous resistance is the minimum possible resistance of the system and corresponds to the resistance during diastole with the coronary bed fully dilated. This component is relatively constant over long periods of time. (*b*) So-called autoregulatory resistance relates to the number and caliber of open channels at the microcirculatory level. Of particular importance are arteriolar smooth muscle tone and capillary density. Autoregulatory resistance is ordinarily 4–5 times larger than basal viscous resistance and can change over a few cardiac cycles in response to changing myocardial metabolic demand. It is traditionally considered the primary mechanism by which coronary flow adjusts to changing demand. Alterations in autoregulatory resistance are mediated primarily by arterioles and/or precapillary sphincters, resulting in changes in total vascular cross-sectional area. The ability of myocardial vessels to vasodilate in response to increased demand is often referred to as "coronary reserve" and is discussed below. The normal magnitude of this reserve is such that flow can increase by a factor of 4–5 without a change in inflow pressure. (*c*) So-called compressive resistance results from the actions on myocardial vessels of local forces resulting from intramyocardial pressure. Compressive resistance varies during a single cardiac cycle, and is especially large during systole. When superimposed on basal viscous and autoregulatory resistance during systole, it reduces flow velocities to ~⅓ of diastolic values.

All three of these functional components of coronary vascular resistance can change temporally, spatially, and/or transmurally. While basal viscous resistance is normally constant over long periods of time, some studies indicate an inherent gradient of vascularity favoring the inner, subendocardial portion of the myocardial wall (15, 16). Compressive resistance changes transmurally as well as temporally. During systole there is an important pressure gradient across the myocardial wall. While the detailed pattern of the gradient is unsettled, intramyocardial pressure is agreed to be near ventricular pressure in the subendocardium and to decrease progressively toward the epicardium (17). As a result, there is relatively little flow to the subendocardium during systole. Compressive resistance is much smaller during diastole, although a similar inner-to-outer gradient is likely at normal preload (18, 19) and is accentuated at elevated preload (19). Absolute levels of diastolic intramyocardial pressure remain controversial, largely because of technical limitations of the various techniques used to estimate them. In considering the various expressions of compressive resistance, it is sometimes unclear whether effects are manifest through a change in P_{ZF} (and therefore driving pressure), a change in vascular cross-sectional area (and therefore resistance), or both.

Autoregulatory resistance changes temporally, transmurally, and spatially. The importance and magnitude of the temporal adjustment to changing myocardial demand has already been commented upon. Transmural and spatial variations are covered in more detail subsequently. The control of autoregulatory resistance is of primary importance in the regulation of coronary blood flow and has been the subject of intense investigation for decades. Mechanisms for adjusting autoregulatory tone can be classified as neurohumoral, metabolic, and myogenic.

Control Mechanisms

Coronary vascular smooth muscle is subject to neurohumoral influences through direct autonomic innervation as well as circulatory vasoactive agents. All studies of neural mechanisms for adjusting coronary resistance have necessarily involved rigid control of myocardial hemodynamics in order to separate direct neural effects from secondary vasodilation induced metabolically as a result of myocardial sympathetic stimulation. In many preparations neural effects were minimized abnormally by anesthesia. On the basis of studies of hearts with intact coronary circulation and isolated segments of coronary vessels, it is acknowledged that direct adrenergic innervation involves both constrictor and dilator mechanisms (14). The former appear to predominate. Experiments in anesthetized and, more recently, conscious animals indicate that an alpha adrenergic mechanism normally exerts a tonic influence on coronary vascular smooth muscle and can serve as a "restraint" on the ability of the coronary bed to vasodilate during stress (20–27). A preliminary report by Vatner and colleagues (28) suggests that an alpha adrenergic mechanism may modulate vascular tone in epicardial coronary arteries as well as at the arteriolar level. Cholinergic regulation of coronary flow has also been demonstrated in experimental animals (29).

Although evidence of alpha adrenergic constrictor influences is more limited in man, a recent study contrasting normal subjects and cardiac allograft recipients (30) suggests that measurable vasoconstrictor tone is present in the normal coronary bed under basal conditions. This vasoconstrictor tone may be involved in the production of angina in some patients with coronary artery disease (31). Coronary artery vasospasm, a common pathophysiologic mechanism in the variant angina syndrome, may also be related to abnormalities in the sympathetic nervous system (32). Maseri et al (33) suggest that coronary spasm is even responsible for myocardial infarction in selected patients with coronary disease. Thus, interest in, and appreciation of, sympathetic neural mechanisms for modulating coronary flow in man is clearly increasing.

Of all factors considered to be involved in the control of autoregulatory resistance, metabolic factors play the largest role. This conclusion is based largely on the close parallel between myocardial O_2 utilization and coronary flow in a myriad of investigations. While numerous agents have, at one time or another, been proposed as "the" metabolic regulator of resistance, the largest body of evidence has been accumulated in favor of adenosine by Berne and his colleagues (34).

Adenosine is a nucleoside composed of ribose and the purine base adenine. It is formed as a result of cleavage of a phosphate group from 5'-adenylic acid (5'-AMP) by the enzyme 5'-nucleotidase, which is located at the myocardial cell membrane. With myocardial hypoxia, there is presumed to be a gradual increase in the concentration of 5'-AMP. This increase provides more substrate for 5'-nucleotidase resulting in increased amounts of adenosine in the interstitial space. The adenosine produces vasodilation by an unexplained action on resistance vessels, resulting in an increase in coronary blood flow. Thus, adenosine responds to a fall in intracellular oxygen tension. Adenosine in the interstitial space is ultimately removed by three processes: conversion to inosine and hypoxanthine in red cells and vascular endothelium; transport back across the myocardial cell membrane with subsequent deamination to inosine or rephosphorylation to 5'-AMP; or removal via increased blood flow. In view of these pathways, it is understandable that adenosine is ordinarily found in very small concentrations (10^{-9} M) in myocardial tissue and coronary sinus blood (34, 35). It is present in increased amounts in ischemic myocardial tissue (35) and in coronary sinus outflow of patients with coronary artery disease during pacing-induced angina pectoris (36).

Prostaglandins are another category of agents that have recently received attention in the metabolic regulation of coronary blood flow. These agents are diversely and potently vasoactive (37 38); regional release increased following experimental coronary artery occlusion (39). This area is still evolving rapidly, however, and the precise role of prostaglandins in the control of coronary flow in man remains unclear. Other potential metabolic regulators of coronary perfusion include oxygen tension, carbon dioxide tension (40, 41), lactic acid, potassium, phosphate, and pH. The careful studies of CO_2 by Case et al (40, 41) are noteworthy, particularly in view of the remarkable elevations in tissue CO_2 reported during experimental ischemia (42).

In summary, available evidence still points more strongly to adenosine as the metabolic regulator of autoregulatory resistance than to any other substance. At the same time, other vasoactive agents may play a role, perhaps through a mechanism such as local modulation of adrenergic transmitter release (26).

The myogenic hypothesis for control of coronary perfusion was originally proposed by Bayliss in 1902 (43). According to this hypothesis, blood vessels are intrinsically able to respond to changes in transmural pressure, i.e. distention of a vessel by an increase in perfusion pressure stimulates contraction of vascular smooth muscle. Similarly, decreases in transmural pressure result in vasodilation. Although it has been difficult to formulate studies capable of defining the importance of myogenic factors in vivo, they seem likely to be involved in the reactive hyperemia following transient coronary artery occlusion (44, 45). They could also conceivably be involved in the control of P_{ZF}.

CORONARY RESERVE AND TRANSMURAL VARIATIONS IN MYOCARDIAL PERFUSION

Overall Coronary Reserve

As mentioned earlier, coronary flow can ordinarily increase several-fold without a change in inflow (aortic) pressure. Commonly referred to as coronary "reserve," this capacity has been related to an inherent ability to decrease coronary vascular resistance several-fold. In Figure 2, this would correspond to increasing the slope of the thick dotted pressure-flow line. The maximum possible increase would correspond to the slope of the line at the left representing maximum vasodilation. While changes in coronary vascular resistance are the major mechanism involved in coronary reserve, changes in P_{ZF} may also play a role, i.e. options for adjusting flow may include changes in position as well as slope of the diastolic pressure-flow relationship. Bellamy reported substantial decreases in P_{ZF} following total inflow occlusion or maximum pharmacologic vasodilation (13).

Our laboratory recently observed that P_{ZF} also varies during modest changes in coronary pressure and flow traditionally attributed entirely to changes in coronary resistance (13a). Decreases in P_{ZF} occur as part of the compensation for effects of arterial stenosis. With mild stenoses, these decreases sometimes allow driving pressure to be maintained at near-normal levels, with preservation of basal perfusion without an appreciable change in resistance. We have also found that P_{ZF} and coronary vascular resistance change in directionally opposite fashion during the early phase of reactive hyperemia following total inflow occlusion (45a). This observation further suggests that P_{ZF} and resistance are controlled by independent mechanisms. When coronary reserve has been exhausted, the coronary bed is presumed to be maximally dilated and the circulation to be operating on the "maximum vasodilation" line illustrated in Figures 1 and 2. The position of this line is determined by P_{ZF}, which is presumably at or near a minimum value in situations in which maximum vasodilation has occurred. The demonstration that the diastolic pressure-flow curve can be shifted to the right by increased preload in the face of maximum vasodilation (19) reminds us that local tissue forces can alter flow independently of changes in vascular resistance and/or inflow pressure.

Figures 1 and 2 are also of interest from the viewpoints of anemia and left ventricular hypertrophy. The anemic patient requires a higher coronary flow to achieve normal O_2 delivery at any given level of myocardial metabolic demand. As outlined by Hoffman (12), the horizontal portion of the overall pressure-flow relationship, i.e. the portion within the autoregulatory range, is therefore elevated. Since the elevated portion intersects the maximum vasodilation line at a higher-than-usual pressure, maximum vasodilation occurs at a higher-than-usual inflow (aortic) pressure. In the case of hypertrophy, proliferation of vascular channels may not always keep pace with the increase in myocardial mass (46). If perfusion is expressed on the basis of flow per unit mass, the slope of the maximum vasodilation line in Figures 1 and 2, which represents maximum possible coronary conductance, is decreased. This again causes maximum vasodilation to occur at a higher-than-usual inflow pressure. Thus, patients with anemia and ventricular hypertrophy are particularly vulnerable to myocardial ischemia during periods of hypotension.

Transmural Myocardial Perfusion

During recent years it has become clear that the principles outlined for overall left ventricular perfusion require important refinement when considering perfusion in different tissue layers. Since mechanical effects of ventricular contraction impede systolic perfusion to a greater degree in the inner layers of the heart than the outer, the inner layers must receive a proportionately greater fraction of their perfusion during diastole. Myocardial O_2 requirements are at least as great in the subendocardium as the subepicardium (17, 47); possible transmural differences in local O_2 extraction (48–51) do not negate the need for the full-cycle level of subendocardial flow at least to equal that in the subepicardium. Diastolic flow to the inner layers of the heart therefore exceeds that to the outer, and mechanisms for achieving this augmentation in diastolic endocardial perfusion continue to be studied intensively. Potential options relate to transmural differences in resistance and driving pressure.

Subendocardial resistance relates to both capillarity and degree of vasodilation. Studies several years ago indicated a greater number of open capillaries in the subendocardium during normal perfusion (52, 53), and it was reasonably suggested (47) that the subendocardium normally uses a portion of its vasodilatory reserve to compensate for the "throttling" effect of systole on coronary inflow. More recent work indicates that minimum possible resistance (during maximum vasodilation) is less in the subendocardium than in the subepicardium (15, 16, 18), i.e. that total vasodilatory reserve is inherently greater in this area. Potential transmural differences in driving pressure also require further investigation. Diastolic driving pressure is pivotal for the subendocardium; a smaller value than in the subepicardium has been presumed on the basis of the normal difference between intraventricular and pericardial diastolic pressures. The recognition that P_{ZF} ordinarily exceeds intraventricular diastolic pressure implies that driving pressures in all transmural layers have been overestimated frequently—particularly at low aortic pressures—and that attention must be focused on the absolute magnitude of P_{ZF} as well as on possible transmural differences. Recent studies by Rouleau and colleagues (18), using pressure-flow data derived from microsphere measurements made during full-cycle perfusion and maximum vasodilation, conclude that transmural differences in P_{ZF} play an important role in transmural flow regulation (with subendocardial P_{ZF} normally exceeding that in the subepicardium). Data from our own laboratory (19), obtained with microsphere measurements made during selective diastolic perfusion and maximum vasodilation, suggest that a directionally similar transmural gradient in P_{ZF} at normal preload is accentuated at elevated preload.

What is the impact of these various factors on transmural perfusion? A convincing body of evidence, recently summarized by Hoffman (12), now indicates that ischemia is ordinarily evident earlier and to a greater degree in the subendocardium than in the subepicardium. Although local perfusion depends on the complex interplay between local resistance and driving pressure (as modulated by autoregulation), coronary reserve is exhausted earlier in the inner layers of the heart under most conditions (18, 54; J. C. Giacomini & F. J. Klocke, unpublished observations). Once reserve has been exhausted, flow at any given inflow pressure becomes critically dependent upon local resistance and the local forces governing "back pressure." Subendocardial ischemia reflects an inadequacy of flow for local O_2 demand despite full utilization of coronary reserve mechanisms.

Attempts to define hemodynamic indices reflecting such an imbalance began with the coronary–left ventricular pressure ratio of Griggs & Nakamura (55). In a widely admired series of studies over the past several years, Hoffman, Buckberg, and colleagues carefully defined advantages and limitations of the DPTI : SPTI ratio (17). DPTI, the diastolic pressure-time index, is the area between the coronary arterial and left ventricular pressure curves during diastole and is considered to reflect O_2 supply. SPTI, the systolic pressure-time index, is the area under the left ventricular pressure curve during systole and is taken to reflect O_2 demand. When the DPTI : SPTI ratio falls below 0.4, coronary reserve is ordinarily exhausted, with local perfusion falling short of O_2 requirements. The ratio has great conceptual attractiveness and practical utility. Nevertheless, specific applications must be made in the context of potential limitations. As recently summarized by Hoffman (12), these limitations relate both to SPTI as an index of local O_2 demand, and to DPTI as an index of local O_2 supply. While effects of anemia can be corrected reasonably easily, effects of transmural variations in minimum possible resistance and driving pressure are more troublesome. Mechanistic interpretations of changes in endocardial-epicardial flow ratios, which can now be measured accurately and reproducibly in experimental animals with radioactive microspheres (56), require consideration of several simultaneously operative factors. In addition to local driving pressures and resistances, these include the absolute levels of endocardial and epicardial flow and diastolic aortic pressure. Decreasing values of the latter assume increasing importance, since they accentuate transmural differences in driving pressure related to transmural differences in P_{ZF}. In the presence of coronary artery disease, the situation is even more complex since diastolic aortic pressure does not reflect coronary arterial pressure downstream to an obstructive lesion. The latter pressure, which is the one needed for determining DPTI, cannot ordinarily be measured in man.

Although of less direct clinical impact, it should also be kept in mind that myocardial perfusion is heterogeneous in a nontransmural as well as a transmural sense. Bassingthwaighte and colleagues (57, 58) characterized the distribution of perfusion, as measured by the radioactive microsphere technique, in 250 segments of the conscious baboon ventricle under a variety of physiological conditions. They report a broad distribution of regional flows (varying from ~0.3 to ~2.2 times the mean), which remains stable during bicycle exercise or hyperthermia. Schanzenbächer & Klocke (59) recently confirmed that regional flows vary similarly within the endocardial, mid-myocardial and epicardial layers of the normal canine ventricle. They also found that heterogeneity of flow within individual layers persisted when average flow for the layer was altered in relation to average flow for the entire ventricle.

SPECIAL PROBLEMS IN ISCHEMIC HEART DISEASE CAUSED BY CORONARY ATHEROSCLEROSIS

Factors Governing the Relationship Between Degree of Stenosis and Flow

The effects of partial coronary artery obstruction on coronary perfusion are complex. A sizable body of information about the resistance to flow offered by stenoses of varying geometry and severity has become available during the past few years (e.g. 60–62). A laboratory model of coronary stenosis has been able to define flow velocity at multiple points across an individual cross section of a vessel, and to develop velocity flow profiles by extending such measurements to sequential cross sections (60). As shown schematically in Figure 3, flow profiles are altered only slightly in the vicinity of a mild stenosis; flow reattaches, so to speak, to the vessel wall immediately downstream of the stenosis. The relationship between pressure drop across the stenosis and flow is therefore linear, and stenosis resistance remains constant as flow increases. With severe stenoses, flow profiles separate from the vascular wall downstream of the stenosis, leading to local turbulence and important energy losses. Pressure drop across the stenosis is greater at all levels of flow; more importantly, pressure drop rises exponentially rather than linearly as flow and/or the degree of stenosis increases. The effective

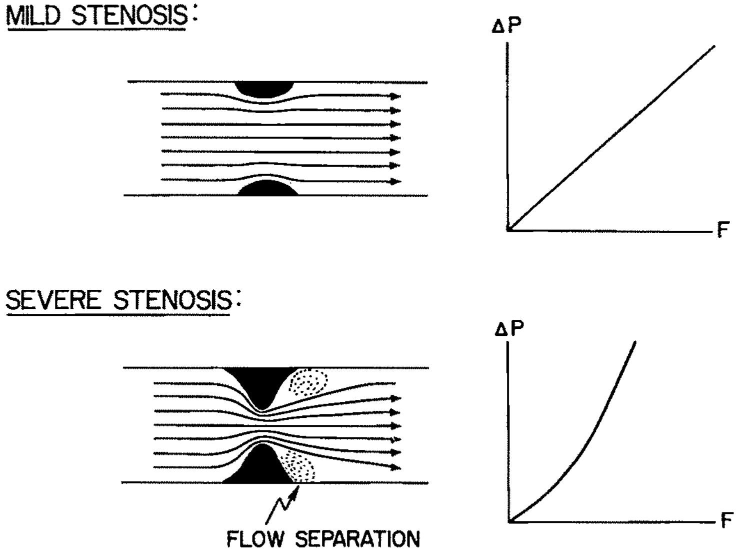


Figure 3 Velocity profiles and pressure-flow relationships for mild and severe stenoses (see text for details): ΔP = pressure drop across the stenosis (only), F = flow.

resistance of the stenosis therefore increases nonlinearly as flow or degree of stenosis increases. Augmentations in the flow increase resistance during situations of increased O_2 demand. Resistance also varies more widely than usual in a stenotic vessel during the cardiac cycle, placing diastolic (and therefore subendocardial) perfusion in particular jeopardy.

In order to examine the in vivo relationship between degree of stenosis and coronary perfusion, the combined effects of the stenosis and the various other factors governing coronary flow must be considered. The left-hand portion of Figure 4 again illustrates that stenosis resistance increases with severity of stenosis. Rather remarkable increments can occur with minimal progression in stenoses obstructing 80% or more of luminal cross-sectional area. The right-hand portion of Figure 4 schematically summarizes the relationship between coronary flow and percentage of stenosis in the intact coronary bed. Basal flow is designated by an arbitrary value of 1.0 and normal coronary reserve potential is represented by the shaded area above this value. The normal situation is represented by 0% stenosis. In this setting, total available coronary reserve is taken as four-fold. As stenosis develops, reserve begins to be compromised but basal flow is preserved. The degree of compromise increases exponentially at the higher degrees of stenosis and reserve is exhausted at ~85% obstruction of arterial cross-sectional area. When the degree of stenosis is in this range, large reductions in flow are produced by even small increments in the degree of obstruction. An

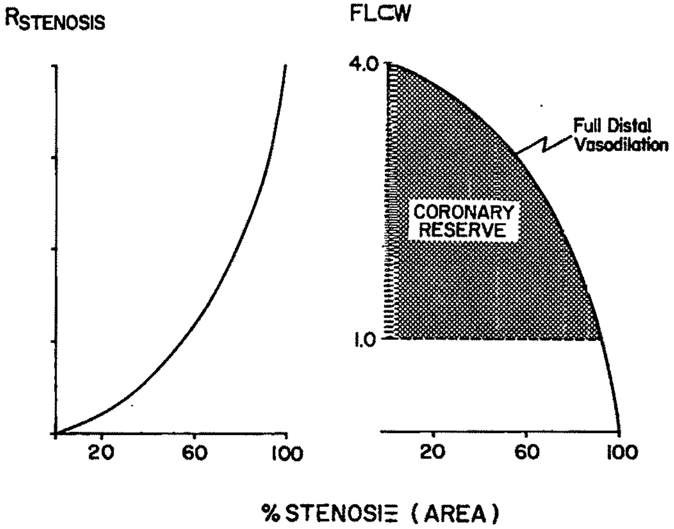


Figure 4 Relationships between stenosis resistance and degree of stenosis, and in vivo flow and degree of stenosis (see text for details).

increase in occluded area from 90 to 95% can potentially reduce flow by a factor of two. These factors underscore the difficulties involved in quantitative inferences about coronary flow from even technically superior arteriograms.

Quantitation of Myocardial Flow in Man

Measurements of myocardial flow in man during the past 2–3 decades have most frequently employed some form of diffusible tracer technique. Until relatively recently, most approaches made no distinction between normally and abnormally perfused areas, but instead purported to give an average value of flow, expressed as ml/min/100 g, for the entire left ventricle. This value was obtained by measuring the rate of entry into or exit from myocardial tissue of an inert tracer. Methodological details varied importantly with different tracers, and conflicting results were frequent in patients with coronary artery disease. It now seems clear that this confusion occurred because localized areas of reduced flow were often incompletely represented in the measurement of average flow (9). In our own laboratories average left ventricular flow has been found to be less under basal conditions in the catheterization laboratory in patients with angina pectoris and double- and triple-vessel disease than in patients with atypical chest pain and normal coronary arteries (63). This finding was surprising in that we (like others) originally thought it likely that differences in flow between normal and coronary patients would be apparent only during stress. One possibility which could not be excluded was that the reduced values of average flow in coronary patients merely reflected inclusion in the flow measurement of areas of scar tissue as well as functioning myocardium.

In recent years, investigators attempting to quantitate flow have appropriately focused on regional rather than overall left ventricular perfusion. Techniques continue most commonly to involve diffusible tracers; tracer exchange is localized by either precordial counting of radioactivity originating from a localized area within the ventricle, or selective sampling of venous outflow from a localized ventricular region. A detailed discussion of different approaches and their methodological requirements is beyond the scope of this paper but has been presented elsewhere (9). Patients with proximal left anterior descending stenoses are particularly suitable for studies employing venous outflow sampling, since great cardiac vein blood originating essentially entirely from myocardium nourished by the anterior descending artery can be sampled in the catheterization laboratory (64). Studies of such patients in our own laboratories (65) indicate that regional flow is decreased consistently under basal conditions with stenoses obstructing 90–100% of luminal cross-sectional area. Regional flow also does not show the usual relationship with double product, almost appearing to be "fixed" at a lower-than-usual value. Cases of 50–75% stenosis have more variable values of regional flow under basal conditions, and may possibly be subdivided on the basis of the flow measurement. Responses of flow to atrial pacing are frequently blunted with these lesser degrees of stenoses, as compatible with a limited coronary reserve.

The frequency with which regional flow has appeared to be limited—at least in regard to usual flow–double-product relationships—in the absence of chest pain, lactate production, etc has again been surprising. Conceivably the limitation is an "artifact" of the inadequacies of systolic pressure as an index of wall stress discussed in the first section. We have been unable to relate it to previous episodes of myocardial necrosis as judged by clinical, enzymatic, electrocardiographic, or angiographic data. Accordingly, we suggested that a chronic reduction in basal regional O_2 demand, perhaps related to altered mechanical performance, should be considered.

Of several recent studies potentially bearing on this point, the most relevant is the Mayo Clinic investigation of anterior wall performance in patients with proximal anterior descending obstruction (66). Peak rates of systolic wall thickening and diastolic wall thinning were decreased systematically under basal conditions in patients with past episodes of ischemia but not infarction. The diastolic abnormality was more severe in patients with 90% or greater obstruction of anterior descending cross-sectional area than in patients with less severe obstruction. These findings are compatible with the hypothesis that a normal balance between O_2 demand and supply can be maintained by local adjustments in regional O_2 demand in the presence of restricted O_2 supply. They and our own data are also consistent with the experimental data summarized above indicating that vasodilatory reserve downstream to an ~85% stenosis can be exhausted under basal conditions.

Potential effects of collateral circulation and physical training on human myocardial perfusion can be mentioned only briefly. While there is no question about the existence and beneficial effect of collaterals in patients with coronary atherosclerosis, the quantitative magnitude of perfusion supplied by these vessels at rest and during stress remains to be defined. It seems reasonable that angiography does not define all functional conduits. Limitations of physiological indices such as retrograde flow and peripheral coronary pressure have been acknowledged carefully by investigators working in the field. Peripheral coronary pressure requires additional review in light of current concepts concerning P_{ZF}. Species differences importantly limit the extrapolation of studies in experimental animals to man. In man as well as the experimental animal, the interpretation of specific findings often requires knowledge of the amount of scar tissue (as opposed to viable myocardium) in the collateralized bed.

Physical training may have several beneficial effects on the coronary circulation. It is well established that myocardial O_2 requirements are less for any given level of exercise in a trained individual. The diminished heart rate is an important factor in this response, while effects of hypertrophy and/or cavity dilation (67) are more difficult to sort out. In addition to reducing O_2 demand, the slower heart rate allows a greater duration of diastolic coronary perfusion. Several studies suggest that myocardial capillary density may be increased by training (68–70). Reported effects of training on collateral perfusion vary and no conclusion seems possible at this time. Wyatt et al (71) suggested that training augments effects of catecholamines on myocardial contractility without a change in intrinsic contractile function. Additional fundamental information about myocardial perfusion and performance during exercise and training is needed, particularly in the light of the evolving concepts of driving pressure and resistance described earlier. Much of this information will have to be obtained in experimental animals subjected to extensive chronic instrumentation. Such studies will need to include observations with and without simulation of coronary stenosis. Even with simulation of local stenosis, extrapolation of findings to man will require considerable caution, since human coronary atherosclerosis commonly has an appreciable longitudinal extent (72–74).

Literature Cited

1. Parmley, W. W., Tyberg, J. V. 1976. Determination of myocardial oxygen demand. *Prog. Cardiol.* 5:19–36
2. Krasnow, N. 1964. Isoproterenol and cardiovascular performance. *Am. J. Med.* 37:514–25
3. Sonnenblick, E. H., Ross, J., Covell, J. W., Kaiser, G. A., Braunwald, E. 1965. Velocity of contraction as a determinant of myocardial oxygen consumption. *Am. J. Physiol.* 209:919–27
4. Graham, T. P. Jr., Covell, J. W., Sonnenblick, E. H., Ross, J. Jr., Braunwald, E. 1968. Control of myocardial oxygen consumption: Relative influence of contractile state and tension development. *J. Clin. Invest.* 47:375–85
5. Jorgensen, C. R., Kitamura, K., Gobel, F. L., Taylor, H. L., Wang, Y. 1971. Long-term precision of the N_2O method for coronary flow during heavy upright exercise. *J. Appl. Physiol.* 30:338–44
6. Kitamura, K., Jorgensen, C. R., Gobel, F. L., Taylor, H. L., Wang, Y. 1972. Hemodynamic correlates of myocardial oxygen consumption during upright exercise. *J. Appl. Physiol.* 32:516–22
7. Jorgensen, C. R., Wang, K., Wang, Y., Gobel, F. L., Nelson, R. R., Taylor, H. 1973. Effect of propranolol on myocardial oxygen consumption and its hemodynamic correlates during upright exercise. *Circulation* 48:1173–82
8. Holmberg, S., Serzysko, W., Varnauskas, E. 1971. Coronary circulation during heavy exercise in control subjects and patients with coronary heart disease. *Acta Med. Scand.* 190:465–80
9. Klocke, F. J. 1976. Coronary blood flow in man. *Prog. Cardiovasc. Dis.* 19: 117–66
10. Wyatt, H. L., Mitchell, J. H. 1974. Influences of physical training on the heart of dogs. *Circ. Res.* 35:883–89
11. Mosher, P., Ross, J., McFate, P. A., Shaw, R. F. 1964. Control of coronary blood flow by an autoregulatory mechanism. *Circ. Res.* 14:250–59
12. Hoffman, J. I. E. 1978. Determinants and prediction of transmural myocardial perfusion. *Circulation* 58:381–91
13. Bellamy, R. E. 1978. Diastolic coronary artery pressure-flow relations in the dog. *Cir. Res.* 43:92–101

13a. Ellis, A. K., Klocke, F. J. 1979. Coronary flow autoregulation with and without changes in coronary vascular resistance. *Circulation* 60(II):259 (Abstr.)

14. Klocke, F. J., Mates, R. E., Copley, D. P., Orlick, A. E. 1976. Physiology of the coronary circulation in health and coronary artery disease. *Prog. Cardiol.* 5:1–17

15. Wüsten, B., Buss, D. D., Deist, H., Schaper, W. 1977. Dilatory capacity of the coronary circulation and its correlation to the arterial vasculature in the canine left ventricle. *Basic Res. Cardiol.* 72:636–50

16. Archie, J. P. Jr. 1978. Minimum left ventricular coronary vascular resistance in dogs. *J. Surg. Res.* 25:21–25

17. Hoffman, J. I. E., Buckberg, G. D. 1976. Transmural variations in myocardial perfusion. *Prog. Cardiol.* 5:37–89

18. Rouleau, J., Boerboom, L. E., Surjadhana, A., Hoffman, J. I. E. 1979. The role of autoregulation and tissue diastolic pressures in the transmural distribution of left ventricular blood flow in anesthetized dogs. *Circ. Res.* In press

19. Ellis, A. K., Klocke, F. J. 1980. Effects of preload on the transmural distribution of perfusion and pressure-flow relationships in the canine coronary vascular bed. *Circ. Res.,* Jan. In press

20. Berne, R. M., DeGeestand, H., Levy, M. N. 1965. Influence of the cardiac nerves on coronary resistance. *Am. J. Physiol.* 208:763–69

21. Feigl, E. O. 1967. Sympathetic control of the coronary circulation. *Circ. Res.* 20:262–71

22. Vatner, S. F., Franklin, D., Van Citters, R. L., Braunwald, E. 1970. Effects of carotid sinus nerve stimulation on the coronary circulation of the conscious dog. *Circ. Res.* 27:11–21

23. Vatner, S. F., Franklin, D., Higgins, C. B., Patrick, T., White, S., Van Citters, R. L. 1971. Coronary dynamics in unrestrained conscious baboons. *Am. J. Physiol.* 221:1396–1401

24. Schwartz, P. J., Stone, H. L. 1977. Tonic influence of the sympathetic nervous system on myocardial reactive hyperemia and on coronary blood flow distribution in dogs. *Circ. Res.* 41:51–58

25. Feigl, E. O. 1975. Control of myocardial oxygen tension by sympathetic coronary vasoconstriction in the dog. *Circ. Res.* 37:88–95

26. Shepherd, J. T. 1978. Alteration in activity of vascular smooth muscle by local modulation of adrenergic transmitter release. (Symposium) *Fed. Proc.* 37:179–211

27. Bassenge, E., Mayer, H. E. 1977. Demonstration of alpha-adrenergic coronary control in different layers of canine myocardium by regional myocardial sympathectomy. *Pflügers Arch.* 372:187–94

28. Vatner, S. F., Pagani, M., Manders, W. T. 1979. Alpha adrenergic constriction of coronary arteries in conscious dogs. *Clin. Res.* 27:503A (Abstr.)

29. Feigl, E. O. 1969. Parasympathetic control of coronary blood flow in dogs. *Circ. Res.* 25:509–19

30. Orlick, A. E., Ricci, D. R., Alderman, E. L., Stinson, E. B., Harrison, D. C. 1978. Effects of alpha adrenergic blockade upon coronary hemodynamics. *J. Clin. Invest.* 62:459–67

31. Mudge, G. H., Grossman, W., Mills, R. M. Jr., Lesch, M., Braunwald, E. 1976. Reflex increase in coronary vascular resistance in patients with ischemic heart disease. *N. Engl. J. Med.* 295:1333–37

32. Ricci, D. R., Orlick, A. E., Cipriano, P. R., Guthaner, D. F., Harrison, D. C. 1979. Altered adrenergic activity in coronary arterial spasm: Insight into mechanism based on study of coronary hemodynamics and the electrocardiogram. *Am. J. Cardiol.* 43:1073-79

33. Maseri, A., L'Abbate, A., Baroldi, G., Chierchia, S., Marzilli, M., Ballestra, A. M., Severi, S., Parodi, O., Bragini, A., Distante, A., Pesola, A. 1978. Coronary vasospasm as a possible cause of myocardial infarction. *N. Engl. J. Med.* 99:1271–77

34. Berne, R. M. 1974. The coronary circulation. In *Mammalian Myocardium,* ed. G. A. Langer, A. Brady, 9:251–81. New York: Wiley. 310 pp.

35. Olsson, R. A. 1970. Changes in content of purine nucleoside in canine myocardium during coronary occlusion. *Circ. Res.* 26:301–6

36. Fox, A. C., Reed, G. E., Glassman, E., Kaltman, A. J., Silk, B. B. 1974. Release of adenosine from human hearts during angina induced by rapid atrial pacing. *J. Clin. Invest.* 53:1447–57

37. Needleman, P. 1976. The synthesis and function of prostaglandins in the heart. *Fed. Proc.* 35:2376–81

38. Ogletree, M. L., Smith, J. B., Lefer, A. M. 1978. Actions of prostaglandins on isolated perfused cat coronary arteries. *Am. J. Physiol.* 235:H400–6

39. Berger, H. J., Zaret, B. L., Speroff, L., Cohen, L. S., Wolfson, S. 1976. Regional cardiac prostaglandin release

during myocardial ischemia in anesthetized dogs. *Circ. Res.* 38:566–71
40. Case, R. B., Greenberg, H. 1976. The response of canine coronary vascular resistance to local alterations in coronary arterial P_{CO_2}. *Circ. Res.* 39:558–67
41. Case, R. B., Felix, A., Wachter, M., Kyriakidis, G., Castellana, F. 1978. Relative effect of CO_2 on canine coronary vascular resistance. *Circ. Res.* 42:410–18
42. Khuri, S. F., Flaherty, J. T., O'Riordan, J. B., Pitt, B., Brawley, R. K., Donahoo, J. S., Gott, V. L. 1975. Changes in intramyocardial ST segment voltage and gas tensions with regional myocardial ischemia in the dog. *Circ. Res.* 37:455–63
43. Bayliss, W. M. 1902. On the local reactions of the arterial wall to changes in internal pressure. *J. Physiol.* 28:220–31
44. Eikens, E., Wilcken, D. E. L. 1974. Reactive hyperemia in the dog heart. Effects of temporarily restricting arterial inflow and of coronary occlusions lasting one and two cardiac cycles. *Circ. Res.* 35:702–12
45. Olsson, R. A. 1975. Myocardial reactive hyperemia. *Circ. Res.* 37:263–70
45a. Canty, J. M., Klocke, F. J. 1979. Beat-to-beat changes in driving pressure and vascular resistance during reactive hyperemia. *Circulation* 60(II):101 (Abstr.)
46. O'Keefe, D. D., Hoffman, J. I. E., Cheitlin, R., O'Neill, M. J., Allard, J. R., Shapkin, E. 1978. Coronary blood flow in experimental canine left ventricular hypertrophy. *Circ. Res.* 43:43–51
47. Moir, T. W. 1972. Subendocardial distribution of coronary blood flow and the effect of antianginal drugs. *Circ. Res.* 30:621–27
48. Gamble, W. J., LaFarge, C. G., Fyler, D. C., Weisul, J., Monroe, R. G. 1974. Regional coronary venous oxygen saturation and myocardial oxygen tension following abrupt changes in ventricular pressure in the isolated dog heart. *Circ. Res.* 34:672–81
49. Monroe, R. G., Gamble, W. J., LaFarge, C. G., Benoualid, H., Weisul, J. 1975. Transmural coronary venous O_2 saturations in normal and isolated hearts. *Am. J. Physiol.* 228:318–24
50. Holtz, J., Grunewald, W. A., Manz, R., vonRestorff, W., Bassenge, E. 1977. Intracapillary hemoglobin oxygen saturation and oxygen consumption in different layers of the left ventricular myocardium. *Pfleugers Arch.* 370:253–58
51. Weiss, H. R., Neubauer, J. A., Lipp, J. A., Sinha, A. K. 1978. Quantitative determination of regional oxygen consumption in the dog heart. *Circ. Res.* 42:394–401
52. Myers, W. W., Honig, C. R. 1964. Number and distribution of capillaries as determinants of myocardial oxygen tension. *Am. J. Physiol.* 207:653–60
53. Weiss, H. R., Winbury, M. M. 1974. Nitroglycerin and chromonar on small-vessel blood content of the ventricular walls. *Am. J. Physiol.* 226:838–43
54. Guyton, R. A., McClenathan, J. H., Michaelis, L. L. 1977b. Evolution of regional ischemia distal to a proximal coronary stenosis: self-propagation of ischemia. *Am. J. Cardiol.* 40:381–92
55. Griggs, D. M. Jr., Nakamura, Y. 1968. Effect of coronary constriction on myocardial distribution of iodoantipyrine-131I. *Am. J. Physiol.* 215:1082–88
56. Heymann, M. A., Payne, B. D., Hoffman, J. I. E., Rudolph, A. M. 1977. Blood flow measurements with radionuclide-labeled particles. *Prog. Cardiovasc. Dis.* 20:55–79
57. Bassingthwaighte, J. B., King, R. B., Hales, J. R. S., Rowell, L. B., Smith, O. A. 1977. Stability of heterogeneity of myocardial blood flow in awake, resting and stressed baboons. *Circulation* 56(III):35 (Abstr.)
58. Bassingthwaighte, J. B. 1977. Physiology and theory of tracer washout techniques for the estimation of myocardial blood flow: Flow estimation from tracer washout. *Prog. Cardiovasc. Dis.* 20:155–89
59. Schanzenbächer, P., Klocke, F. J. 1980. Inert gas measurements of myocardial perfusion in the presence of heterogeneous flow documented by microspheres. *Circulation,* March. In press
60. Mates, R. E., Gupta, R. L., Bell, A. C., Klocke, F. J. 1978. Fluid dynamics of coronary artery stenosis. *Circ. Res.* 42:152–62
61. Gould, K. L. 1978. Pressure-flow characteristics of coronary stenoses in unsedated dogs at rest and during coronary vasodilation. *Circ. Res.* 43:242–53
62. Lipscomb, K., Hooten, S. 1978. Effect of stenotic dimensions and blood flow on the hemodynamic significance of model coronary arterial stenoses. *Am. J. Cardiol.* 42:781–92
63. Klocke, F. J., Bunnell, I. L., Greene, D. G., Wittenberg, S. M., Visco, J. P. 1974. Average coronary blood flow per unit weight of left ventricle in patients with

and without coronary artery disease. *Circulation* 50:547–59

64. Nakazawa, H. K., Roberts, D. L., Klocke, F. J. 1978. Quantitation of anterior descending vs. circumflex venous drainage in the canine great cardiac vein and coronary sinus. *Am. J. Physiol.* 234:H163–66
65. Klocke, F. J., Greene, D. G., Bunnell, I. L., Roberts, D. L., Dashkoff, N., Arani, D. T. 1979. Relationship between degree of stenosis and reductions in regional myocardial blood flow in coronary artery disease. *Clin. Res.* 27:502A (Abstr.)
66. St. John Sutton, M. G., Frye, R. L., Smith, H. C., Chesbro, J. H., Ritman, E. L. 1978. Relation between left coronary artery stenosis and regional left ventricular function. *Circulation* 58: 491–97
67. Wyatt, H. L., Mitchell, J. H. 1974. Influences of physical training on the heart of dogs. *Circ. Res.* 35:883–89
68. Wyatt, H. L., Mitchell, J. H. 1977. Influences of physical conditioning and deconditioning upon the coronary vasculature of dogs. *Am. J. Cardiol.* 39:262 (Abstr.)
69. McElroy, C. L., Gissen, S. A., Fishbein, M. C. 1978. Exercise-induced reduction in myocardial infarct size after coronary artery occlusion in the rat. *Circulation* 57:958–62
70. Leon, A. S., Bloor, C. M. 1968. Effects of exercise and its cessation on the heart and its blood supply. *J. Appl. Physiol.* 24:485–90
71. Wyatt, H. L., Chuck, L., Rabinowitz, E., Tyberg, J. V., Parmley, W. W. 1978. Enhanced cardiac response to catecholamines in physically trained cats. *Am. J. Physiol.* 234:H608–13
72. Jones, A. A., Roberts, W. C. 1978. Quantitation of coronary artery narrowing in sudden coronary death. *Circulation* 58:II–133 (No. 516) (Abstr.)
73. Arnett, E. N., Isner, J. M., Redwood, D. R., Kent, K., Baker, W. P., Ackerstein, H., Roberts, W. C. 1979. Underestimation by angiography of critical coronary arterial narrowing in life: Comparison with degree of narrowing assessed by quantitative histological examination at necropsy. *Am. J. Cardiol.* 43:343(Abstr.)
74. Virmani, R., Roberts, W. C. 1979. Quantitation of coronary artery narrowing in clinically-isolated angina pectoris. *Am. J. Cardiol.* 43:343 (Abstr.)

Ann. Rev. Med. 1980. 31:509–40
Copyright © 1980 by Annual Reviews Inc. All rights reserved

CONTROVERSIES IN PLATELET TRANSFUSION THERAPY[1]

❖7376

Sherrill J. Slichter, M.D.

Puget Sound Blood Center and the Department of Medicine,
University of Washington School of Medicine, Seattle, Washington 98104

INTRODUCTION

As early as 1910 Duke (1) reported the efficacy of transfusing platelets in patients with bleeding disorders. However, it was not until the 1950s that platelet transfusions were investigated in a systematic fashion. The early workers in this field, Dillard, Brecher & Cronkite (2), Stefanini et al (3, 4), and Hirsch & Gardner (5), used platelet transfusions primarily as a diagnostic tool to evaluate mechanisms of thrombocytopenia rather than for therapeutic purposes. The advent of plastic bags for blood collection in the 1960s made it feasible to centrifuge blood so that useful concentrates of platelets could be prepared. As a consequence, platelet transfusions were adopted more frequently in the management of thrombocytopenic patients.

In 1964, investigators at the National Cancer Institute (NCI) described the value of platelet transfusions in the treatment of leukemia, where hemorrhage was considered to be the major cause of death in 52% of 414 acute leukemia patients studied from 1954 to 1963 (6). During 1954 through 1959, before platelet preparations were available for transfusion, fatal hemorrhage was observed in 67% of these cases. But in the 3-year period between 1960 and 1963, only 37% of such patients died of hemorrhage. In 1968 Han et al (7) reported the results of autopsy in 57 cases of acute leukemia. Thirty of these patients had died before platelet tranfusions were available, and 65% of these showed major hemorrhage as the proximate cause of death. However, only 15% of the 27 patients for whom platelet transfusions were available died of major bleeding. In a nonrandomized

---

[1] This work was supported by a National Heart, Lung, and Blood Institute Research and Demonstration Center Grant HL-17265.

0066-4219/80/0401-0509$01.00

retrospective study of children with acute leukemia treated with repeated platelet transfusions, fewer bleeding days were recorded than in untreated patients (8). Another report showed that patients given prophylactic platelet transfusions for platelet counts of less than 25,000/μl had a reduction in overall bleeding from 56 to 7%, while minor bleeding episodes fell from 44 to 4% (9). In the only prospective controlled study, 18 patients with acute leukemia and platelet counts less than 30,000/μl were randomized to receive either platelets or platelet-poor plasma (10). Eight of 9 in the plasma-treated group bled, as compared with 4 of 12 in the platelet-treated group. Fever was asssociated with substantial bleeding in 10 of 13 patients. These early studies established the value of platelet transfusions and subsequent studies have further defined their preparation, storage, and use.

Certain general principles of platelet transfusion therapy are widely accepted on the basis of scientific data and clinical practice. These principles are discussed in the next section. In subsequent sections the areas of platelet transfusion management that remain controversial are discussed, including methods of maintaining platelet viability and function during storage, indications for platelet transfusion, factors affecting platelet alloimmunization and techniques for selecting compatible platelet donors. Because the latter subjects remain controversial, additional scientific data will be required to develop firm guidelines.

GENERALLY ACCEPTED PRINCIPLES OF PLATELET TRANSFUSION THERAPY

Collection and Preparation of Platelets for Transfusion

The first platelet transfusions were direct transfusions of fresh whole blood either from normal donors or preferably from patients with polycythemia vera and associated thrombocytosis (3–5). Platelet-rich plasma preparations were subsequently administered when centrifugation techniques permitted this advance. Attempts to prepare platelet concentrates were originally thwarted by inability to resuspend the platelets after they had been centrifuged into a platelet button. This problem was solved by either preparing the platelet concentrate from ACD acidified platelet-rich plasma (11) or allowing nonacidified ACD and CPD anticoagulated platelet concentrates to remain undisturbed at room temperature for 1–2 hours before resuspension (12). Acidification to pH 6.5 inhibits platelet aggregation during centrifugation; after centrifugation, a waiting period before resuspension permits spontaneous disaggregation. Either of these two techniques of platelet resuspension when used with appropriate centrifugation force and time, allows the concentration of approximately 85% of the whole blood platelets into a small volume while maintaining post-tranfusion platelet viability and function (13). This preparation of individual platelet units in a small volume makes it possible to pool several concentrates into an appropriate transfusion dose without producing the volume overload that can occur with platelet-rich plasma or whole blood administration. Furthermore, when platelet concentrates are prepared, the remaining plasma and red cells can be used for other transfusion requirements. Since platelet viability is not maintained under the conditions of whole blood storage at 4°C (14–16) refrigerated blood cannot be used as a source of viable platelets.

Platelet concentrates can either be prepared from units of routinely donated whole blood, or from a single donor. In the latter case, the donor's blood flows into a spinning centrifuge bowl and after the platelet fraction is diverted into a collection bag, the remaining plasma and red cells are returned to the donor. This process is generally repeated until the equivalent of 4–6 units of platelets have been harvested. Both the viability and function of platelets collected by this technique are normal (17), but the procedure should be confined to preparing platelets from a compatible donor for an alloimmunized thrombocytopenic patient.

Selection of Thrombocytopenic Recipients

Several pathologic states may result in thrombocytopenia. The dosage, frequency of administration, and utility of platelet transfusions all depend on underlying etiology.

THROMBOCYTOPENIA SECONDARY TO FAILURE OF MARROW PLATELET PRODUCTION Decreased megakaryocytopoiesis may be congenital, idiopathic, or secondary to radiation, chemicals, insecticides, drugs, viruses, or marrow replacement (18). Patients with this form of thrombocytopenia are the best candidates for platelet transfusions since the increment and survival of transfused platelets should be close to normal. In 23 patients with production-related thrombocytopenia, we found that platelet recovery was 56 ± 3% of the number transfused, and survival was 5.2 ± 1.1 days. These results were interpreted as modestly reduced since normal subjects had 65 ± 5% recoveries and 9.6 ± 0.6 day survivals (Figure 1). To exclude alloimmune destruction, the patients selected for this study either had autologous platelet survivals measured, received platelets from an identical twin, or, if transfused with random donor platelets, they had never been previously transfused or pregnant. The modest reduction in platelet survival observed in these thrombocytopenic patients has been previously reported (19–21) but the mechanism for the shortened platelet survival has never been established. The best hypothesis may be that platelets normally interact in some fashion with the vascular endothelium (22–27). When the platelet count is normal, the percentage of platelets involved in this process

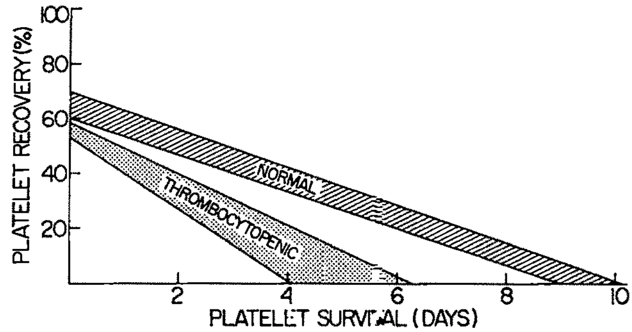


Figure 1 Platelet recovery and survival in normal and in thrombocytopenic patients. In 30 normal individuals initial increment (i.e. recovery) of ^{51}Cr-labeled autologous platelets was 65 ± 5% (S.D.) and survival, 9.6 ± 0.6 days. In 23 patients with production-related thrombocytopenia and platelet counts of <70,000/μl, autologous or allogenic platelet recovery was 56 ± 3% and survival was 5.2 ± 1.1 days. To exclude alloimmune destruction, the thrombocytopenic patients selected to receive donor platelets either had never been previously transfused or pregnant or were given platelets from an identical twin.

may be so small that it cannot be measured. However, in thrombocytopenic patients this vascular consumption may be proportionately much larger and thus readily detected. In such patients platelet transfusions are usually required about twice a week to maintain circulating platelet levels adequate for hemostasis.

THROMBOCYTOPENIA SECONDARY TO SPLENIC POOLING Normally about one third of the platelets produced by the bone marrow are pooled in the spleen (28). In hypersplenic individuals the platelet level is usually in the range of 30,000–80,000/μl, so platelet therapy is not a common requirement. However, if platelet transfusions are needed in these individuals, recovery (i.e. incremental response following transfusion) is reduced proportionately to the size of the spleen. Thus, these patients need more platelets than normosplenic individuals to achieve the same post-transfusion platelet count. Fortunately, platelet survival is not altered (Figure 2). If platelets are used to control operative bleeding during splenectomy, maximum response is achieved by clamping the splenic pedicle before the platelets are transfused.

THROMBOCYTOPENIA SECONDARY TO INCREASED DESTRUCTION The difficulty with platelet transfusions in this group of patients is that the factors affecting the survival of the patients' own platelets are almost always extrinsic to the platelet and therefore will cause at least as much destruction

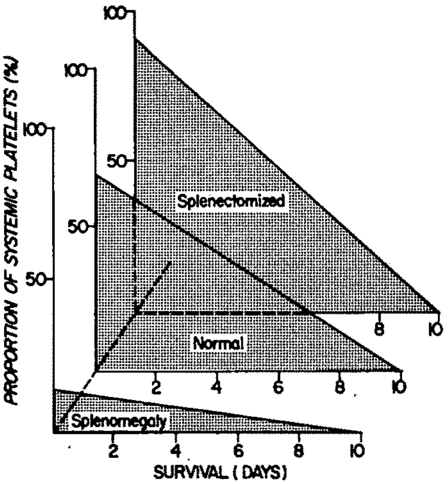


Figure 2 The effect of spleen size on the circulating platelet count. Normally about one third of the platelets produced by the bone marrow are pooled in the spleen. In patients with hypersplenism the proportion of circulating platelets is reduced while platelet survival is unchanged. In asplenic individuals approximately 90% of the platelets circulate. Similarly, the incremental platelet count (i.e. recovery) following platelet transfusions is directly related to the size of the recipient's spleen. (From Ref. 28.)

of donor platelets. The frequency of platelet transfusions in this group has to be increased, and in certain situations it becomes impossible to sustain an effective increment. However, some of these patients do not require increased platelet levels because their own platelets are young and have increased hemostatic effectiveness.

Autoimmune thrombocytopenic purpura (ITP) In patients with severe ITP, transfused platelets survive for only a few hours (29), so they should not be transfused unless the patient has life-threatening bleeding. Such bleeding, even with very low platelet counts, is rare in ITP because the compensatory marrow response is associated with a reduction in megakaryocyte maturation time and release of young hyperfunctional platelets. The presence of such platelets is most clearly shown in patients with ITP by measurements of the relationship of bleeding time to platelet number (Figure 3). Specific treatment of the underlying disease with corticosteroids usually reduces the accelerated platelet destruction and allows the compensatory marrow platelet production to exceed the rate of removal, so the peripheral platelet count rises. This type of specific therapy should be used as the primary treatment rather than attempting to maintain platelet counts by transfusion.

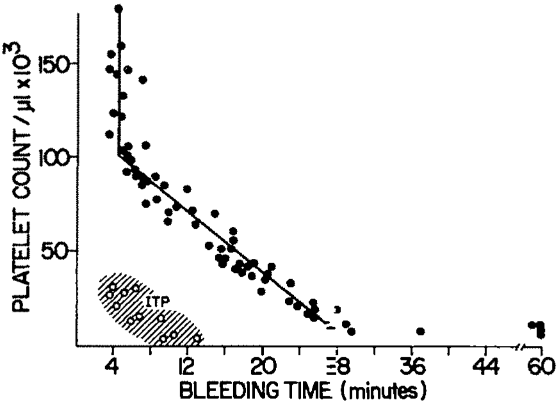


Figure 3 Relationship between bleeding time and platelet count in production-related thrombocytopenia versus patients with autoimmune thrombocytopenia (ITP). The template bleeding time shows a direct inverse relationship to platelet counts between 100,000 and 10,000/µl in patients with production-related thrombocytopenia (*closed circles*). However, in patients with ITP (*open circles*), the bleeding time is disproportionately reduced indicating the presence of hyperfunctional platelets. Thus, patients with ITP have less tendency to bleed at any given platelet count than those with failure of marrow platelet production. (From Ref. 51.)

Disseminated intravascular coagulation The common denominator in this group of disorders is extensive tissue injury with resultant increased utilization of platelets as well as other plasma coagulation factors (30). Situations in which there is sufficient consumption to require platelet transfusions include massive trauma or burns (31), bacteremia, leukemia, and extensive surgery, particularly open heart surgery where there is an additional loss of platelets that adhere to artificial surfaces in the extracorporeal circuit. Platelet transfusions in these situations are given only for supportive reasons to control or prevent bleeding until the underlying disease process resolves with specific treatment.

Some of these conditions require special mention because of their frequency, severity, or unusual demand for platelet support. After extensive trauma, burns, or surgery, patients often require massive blood transfusions. The combination of increased peripheral platelet utilization plus volume replacement with stored blood lacking viable platelets may produce severe thrombocytopenia. After the administration of approximately 16 units of blood, thrombocytopenia (platelet count 100,000/µl or less) may contribute to bleeding, and platelet transfusions should be considered (Figure 4) (32). Patients undergoing open heart surgery may have not only thrombocytopenia but also platelet dysfunction related to cardiopulmonary bypass and hypothermia. Thus, cardiac patients who bleed post-operatively in spite of platelet counts greater than 100,000/µl may have continued

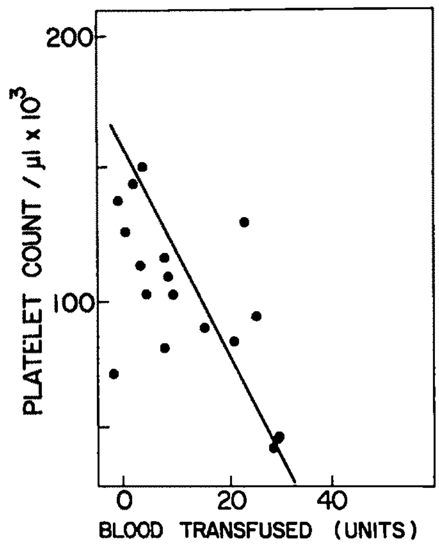


Figure 4 Massive blood transfusions and dilutional thrombocytopenia. In patients with extensive transfusion requirements the platelet count is inversely related to the number of units of blood transfused. With platelet counts of <100,000/µl and a disrupted vascular system, platelet transfusions may be required to permit platelet plug formation. (From Ref. 32.)

platelet dysfunction off bypass. These patients should benefit from the transfusion of functional platelets (33).

Patients with acute leukemia present very difficult support problems. Not only do they have the most extensive platelet consumption of all the malignancies (Figure 5) (34), but marrow invasion with leukemic cells plus marrow-ablative chemotherapy results in failure of compensatory marrow platelet production. Furthermore, compromised marrow production involves granulocytes as well, and the associated septicemia further reduces the survival of transfused platelets (30). Platelet support requirements in leukemia are often more excessive and prolonged than in any other clinical condition with the exception of chronic refractory marrow aplasia. Alloimmunization is a frequent occurrence, with the attendent problems of finding compatible platelet donors.

Drug-related thrombocytopenias Quinidine is the most common cause of drug-induced immune thrombocytopenia (35). Other drug-related platelet antibodies have also been documented (36), and all of these are characteristically self-limited with discontinuation of the drug. When thrombocytopenia is so severe that platelet transfusions are required to control bleeding, donor platelets are also rapidly destroyed so that transfusions may have to be provided once or twice a day until drug elimination.

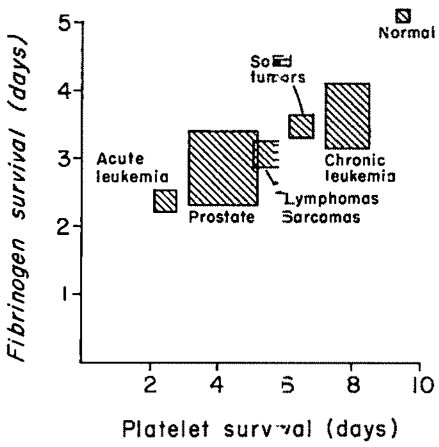


Figure 5 Platelet survival is reduced in malignant diseases. In patients with metastatic malignancies, the combined consumption of platelets and fibrinogen is most severe in the acute leukemias. (From Ref. 34.)

Post-transfusion or neonatal isoimmune thrombocytopenia Several platelet-specific antigen systems have been identified (37). The development of antibodies to these platelet-specific antigens may lead to thrombocytopenia requiring the transfusion of platelets that do not have the offending antigen. Such problems usually are related to the PLA^{-1} antigen, which is carried by 97% of the population. In two clinical settings difficulties may develop in PLA^{-1} negative individuals who have developed anti-PLA^{-1} antibodies, usually from prior exposure to PLA^{-1} positive platelets by transfusion or pregnancy. Rarely, a PLA^{-1} negative individual given a transfusion containing platelets from a PLA^{-1} positive individual develops severe acute thrombocytopenia 3–5 days later. This event is associated with an anamnestic recall of the antibody, which then, by an unknown mechanism, destroys not only the PLA^{-1} positive transfused platelets but also the autologous PLA^{-1} negative platelets. Management consists of transfusing PLA^{-1} negative donor platelets or performing an exchange transfusion with plasma or whole blood to remove the antibody (38). In addition, a PLA^{-1} negative pregnant woman may be sensitized by the PLA^{-1} positive platelets of her unborn child. This IgG antibody can cross the placenta and cause severe thrombocytopenia in the baby. However, since the mother is PLA^{-1} negative, her platelets can be used to provide the baby with compatible platelets just as in hemolytic disease of the newborn. When this procedure is followed, it may be important to wash the platelets free of plasma before transfusion so that additional antibody is not transfused (39).

CONGENITAL DISORDERS OF PLATELET NUMBER OR FUNCTION

The majority of patients with hereditary thrombocytopenia do not have very low platelet counts and, since ineffective platelet production is the usual underlying mechanism, the response to platelet transfusions, if needed, is good (18). The use of transfusions should be limited to the control of operative or traumatic bleeding, since these chronically thrombocytopenic patients are always subject to alloimmunization. Similar recommendations are made for patients with thrombocytopathic conditions causing significant platelet dysfunction (40).

Requirements for Documenting Efficacy of Transfused Platelets

Determining the clinical effect of transfused platelets in a bleeding patient is often difficult because petechiae and purpura take several days to disappear, and bleeding may resolve even without platelet transfusions. To circumvent these problems in human subjects, thrombocytopenic models were developed in animals by total body irradiation or platelet antiserum. The effects of platelet transfusions were then evaluated directly by measuring red blood cell loss into the lymphatic system as well as changes in hematocrit and bleeding times. These studies clearly showed the necessity of viable platelets to control bleeding manifestations, since disrupted, lyophilyzed, or improperly frozen platelets as well as membrane phospholipid fractions were ineffective substitutes (41–44). These initial observations were subsequently confirmed in thrombocytopenic patients (45).

Given that viable platelets are required, the accepted indicators of effectiveness are the platelet increment and survival after transfusion. These values are determined by measuring the platelet count before transfusion and then at 1 and 4 hours and daily after transfusion. Survival measurements determine the required frequency of transfusion while the initial platelet increment correlates most closely with the immediate clinical response (8, 18, 46). For example, in a group of leukemic patients, major vessel bleeding (melena, epistaxis, hematuria, oral mucosal and hematemesis) was controlled 33% of the time (8/24 patients) when the post-transfusion increment was $<20{,}000/\mu l$, 67% of the time when the count was between 21,000 and $40{,}000/\mu l$ (12/18), and 83% of the time with platelet counts greater than $40{,}000/\mu l$ (13/16). In thrombocytopenic recipients platelet increments can usually be expected to be 56 ± 3% of the number of platelets transfused (18). Causes of reduced increments are hypersplenism, where the reduction is correlated with the size of the spleen (28); alloimmunization (47); and fever or infection. In the latter instances, response may be decreased by 20–40% (8, 9, 48) conceivably as a direct effect of endotoxin or bacteria on the platelets (49). Following 131 transfusions into 28 patients (usual donor was a parent), recoveries were 39% if the patient was afebrile and 21% with temperatures greater than 38.5 ($P < 0.01$). During sepsis recovery was reduced to only 15%. Values for platelet survival for the respective groups was 60% of the initial value on day 2 for noninfected patients, 25% for fever greater than 38.5°C, and 12% for sepsis. Alloimmunization is also associated with reduced platelet survival while splenomegaly is not. Whether fever without infection reduces platelet increments and survivals is not clear, since negative cultures in a febrile patient do not exclude infection.

In addition to viability measurements, the function of the transfused cells is important in the clinical response. Unfortunately, in vitro tests of platelet function (aggregation, adhesion, PF3 release, serotonin uptake or release, and hypotonic shock) may not predict the in vivo function of platelets (50). The most reliable in vivo platelet function test is the standardized template bleeding time (51).

CONTROVERSIAL ASPECTS OF PLATELET TRANSFUSION THERAPY

Having identified the best methods of preparing platelets for transfusion, the expected response, and how it may vary depending on the clinical condition of the patient, several aspects of platelet support still remain controversial because of conflicting or incomplete data.

Platelet Storage

It has been known for over 15 years that storing platelets at 4°C for more than 24 hours results in significant loss in platelet viability (16). By testing platelets stored at several temperatures, it was determined that 22°C is optimum for platelet recovery and survival (52). The better viability of platelets stored at 22°C compared to those stored at 4°C has been confirmed by several investigators (Figure 6) (52–56). However, of equal importance to maintaining platelet viability during storage is the preservation of hemostatic effectiveness. The in vivo functional integrity of 22°C vs 4°C stored platelets has been an area of major dispute.

When the relationship between post-transfusion bleeding time and platelet count was determined for platelets stored at 4°C and 22°C for 24 and 72 hours, it was shown that at 22° platelets maintained their functional integrity, while at 4°C platelets became dysfunctional, except for occasional transient improvements in the bleeding time (Figure 7) (53). This finding was not supported by the first study of post-storage platelet function by Aster and his co-workers, who showed better function by 4°C- than by 22°C-stored platelets (56). However, a third of their patients had no im-

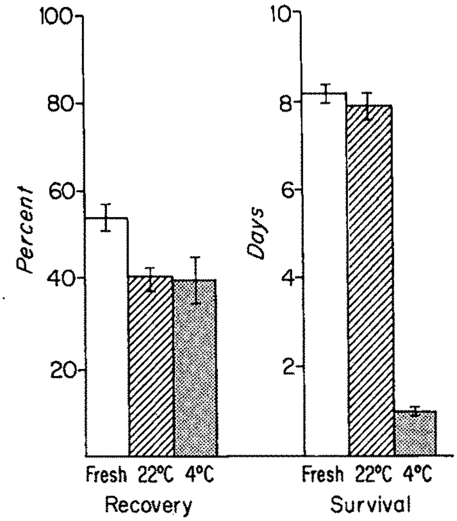
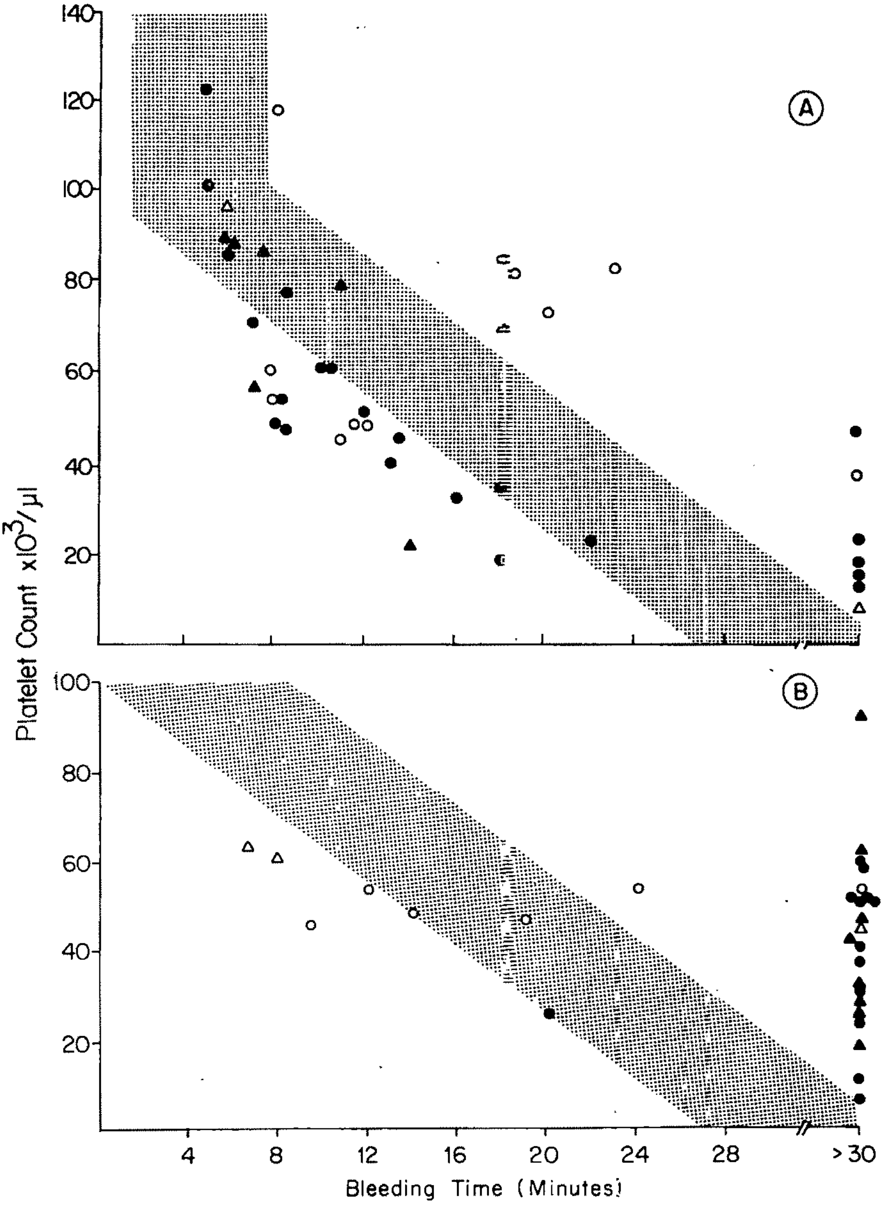


Figure 6 Viability of platelet concentrates stored at 22°C vs 4°C. Platelet concentrates were prepared from normal volunteer donors, and after ^{51}Cr labeling they were transfused immediately or after storage for 72 hours at either 22°C or 4°C. Recovery and survival of the autologous transfused platelets are shown (height of column, mean; bar, ±1 S.E.). (From Ref. 53.)

provement in bleeding time even with fresh platelet transfusions, and, after storage at room temperature, half of the concentrates had a pH below 6.0 because of inadequate plasma volume (<30 ml). Platelet concentrates with a post-storage PH of 6.0 or below are known to be nonviable and therefore are not expected to provide hemostasis. A more recent study was performed by Filip & Aster (57) with platelets stored at room temperature and maintained in 50 ml of plasma. Their results showed that although platelet concentrates stored for 72 hours at 4°C maintained some function during the first 4 hours after transfusion, these platelets were clearly dysfunctional by 24 hours. Conversely, the platelets stored for 72 hours at 22°C were consistently able to shorten the recipient's bleeding time.

Another cause of the controversy over platelet function post-storage was related to the use of normal volunteers given doses of aspirin rather than thrombocytopenic patients as the test subjects. Normal volunteers who had ingested aspirin 24 hours earlier were transfused with platelets, and bleeding times were measured before and after transfusion to determine whether bleeding times prolonged by aspirin could be corrected (58). O'Brien (59) showed that if nonaspirinized platelets constituted at least 10% of the total number of platelets being evaluated, platelet function as measured by aggregation was normal. To achieve this 10% platelet level in vivo, Handin & Valeri (58) gave 4 units of pooled homologous fresh platelet concentrates


A
B
140
120
100
80
60
40
20
Platelet Count x10^3/µl
4
8
12
16
20
24
28
>30
Bleeding Time (Minutes)

to the aspirin-treated normal volunteers and found that the bleeding time was corrected to normal within 2 hours of transfusion. Autologous platelet concentrates obtained by plateletpheresis before aspirin ingestion were then stored for 24 or 48 hours at room temperature. Two hours after transfusion of the stored pooled concentrates into the aspirinized donors, there was no correction of the post-transfusion bleeding time. However, a correction was found 24 hours post-transfusion. This finding indicated a loss of platelet function during storage but in vivo reversability of this dysfunction with time. A subsequent study by Valeri (60) indicated that interpreting platelet function by using this aspirinized normal volunteer model presented difficulty. A single platelet concentrate transfused into aspirinized normal volunteers did not achieve the 10% nonaspirinized platelet level of the prior study, and therefore no correction of the bleeding time occurred with either fresh or room temperature-stored platelets. However, correction did occur following the transfusion of one unit of 4°C-stored platelets; this suggests that there is some in vitro activation of the platelets during 4°C storage, or alternatively, that this model provides artifactual data on platelet function. Indeed, Kevy et al (61) have shown that the aspirinized bleeding time is improved not only after administration of 4°C-stored platelets but also with the infusion of the supernatant plasma removed from the 4°C-stored platelets. Thus it appears that the aspirinized normal volunteer model does not accurately reflect the results of platelet transfusions given to thrombocytopenic patients.

In summary, cold-stored platelets may sometimes be able to provide transient hemostasis, but their routine use for anything other than emergency treatment is not in the best interest of the patient since they do not maintain either function or viability as well as 22°C-stored cells. In addition, the need for more frequent transfusions because of the shortened survival of 4°C-stored platelets would be expected to result in a higher percentage of alloimmunized patients.

Figure 7 Function of platelets after storage at 22°C and 4°C. Bleeding times and platelet counts in aplastic thrombocytopenic patients following transfusion of platelets stored at 22°C (*A*) and 4°C (*B*). Pre-transfusion bleeding times were greater than 60 minutes and platelet counts were less than 10,000 per μl in all cases. (*A*) 22°C: data on four patients following nine transfusions (two stored for 24 hours and seven stored for 72 hours). (*B*) 4°C: data on two patients after eight transfusions (three stored for 24 hours and five stored for 72 hours). Platelets are functional in every case within 2½ hours post-infusion of 22°C stored platelets and remain functional (*A*) while only a single measurement showed any function after 2½ hours post-transfusion of 4°C stored platelets (*B*). Triangles = 24-hour storage; circles = 72-hour storage. Open symbols = measurement between 5 minutes and 2½ hours post-transfusion; closed symbols = after 2½ hours. Shaded area represents 95% confidence limits of relationship between bleeding time and platelet count in untransfused aplastic thrombocytopenic patients. (From Ref. 53.)

Prophylactic vs Therapeutic Platelet Transfusions

In patients who are expected to be thrombocytopenic for extended periods (aplastic, leukemic, or other patients with malignancy undergoing extensive chemo- or radiotherapy) it is important to consider platelet transfusion schedules very carefully since repeated transfusion leads to alloimmunization (62). It would be relatively easy to determine when to provide platelet transfusions if there were a direct relationship between platelet count and/or function and bleeding manifestations. Several studies have attempted to define such relationships. A platelet count of at least 20,000/μl has been widely accepted as the required concentration based on a study by Gaydos et al (63). However, even the original authors indicated that they could not determine a threshold platelet level based on their studies (Figure 8). Grossly visible hemorrhage (hematuria, melena, and hematemesis) was most frequent (10–30% of the days) at platelet counts of less than 5000/μl. Patients whose platelet counts were between 5000 and 100,000/μl had little difference in bleeding (8 and 4%, respectively.) Sixteen of the 92 patients analyzed died of intracerebral bleeding. Half of these had blastic crisis in CML, and autopsy revealed leukostasis and leukemic nodules. In this group, platelet levels were relatively high at the time of hemorrhage (the median being 10,000/μl) and only 1 patient had a platelet count below 5000/μl. By contrast, in patients without blastic crisis, only 1 of the 8 had a platelet count higher than 5000/μl and that count was less than 10,000/μl. A level of 5000/μl is in agreement with stool blood loss studies as a measure of critical platelet levels (Figure 9). The minor forms of hemorrhage such as petechiae, ecchymoses, and epistaxis occurred much more frequently than the more serious forms at all platelet levels (Figure 10) (63).

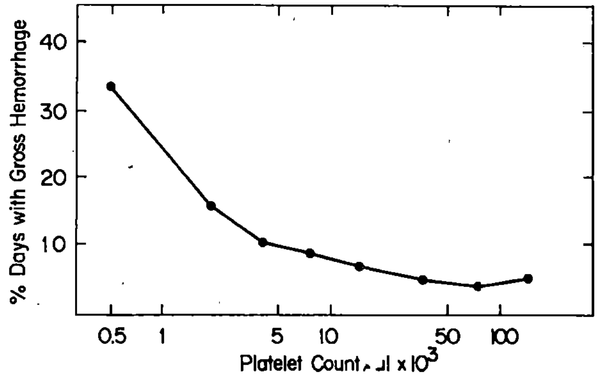


Figure 8 Clinical bleeding as a manifestation of platelet count. In 92 thrombocytopenic leukemic patients grossly visible hemorrhage (hematuria, melena, hematemesis) expressed as a fraction of total bleeding days according to platelet count. (From Ref. 63.)

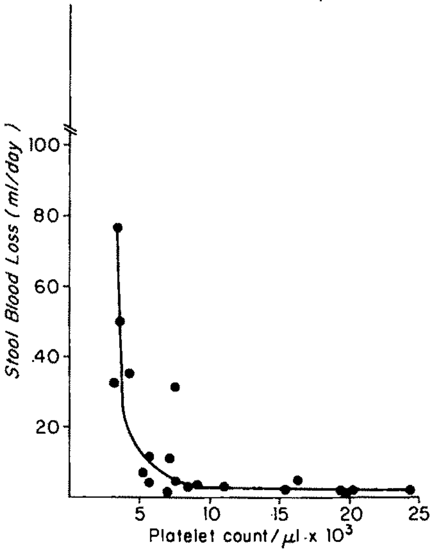


Figure 9 Stool blood loss as a measure of thrombocytopenic bleeding. When stool blood loss (expressed as ml of blood per day) was determined in 20 aplastic thrombocytopenic patients (*closed circles*), blood loss was less than 5 ml per day at platelet counts greater than 10,000/ µl. At platelet counts between 5000 and 10,000/µl, blood loss averaged 9 ± 7 ml per day. At levels less than 5000/µl blood loss was markedly elevated at 50 ± 20 ml per day. (From Ref. 18.)

Similar observations were made in another study of bleeding in children with acute leukemia. Minor bleeding episodes in skin, mucous membranes, microscopic hematuria, guaiac positive stools, and epistaxis were found in approximately 50% of individuals with platelet counts between 0 and 40,000/µl but there was no correlation between frequency of bleeding and absolute counts within that range (64). However, serious bleeding (gross bleeding from GI or GU tract) occurred in 26% of the 62 patients with platelet counts between 0 and 10,000/µl, 10% between 10,000 and 20,000/ µl and 4–5% between 20,000 and 40,000/µl.

Bleeding at levels higher than 5000/µl is most likely related to platelet dysfunction associated with underlying disease (65, 66), fibrinolysis (67), drugs (68–70), or fever/infection (71). In the clinical management of patients with thrombocytopenia it is particularly important to recognize the influence on platelet function of aspirin-containing compounds and semisynthetic penicillins. Although the former can be avoided, the latter drugs often have to be given because of the frequency of low white counts and concurrent infections in patients with thrombocytopenia.

Fever and/or infection may also contribute directly to the bleeding tendency in thrombocytopenic patients although this issue has not been ade-

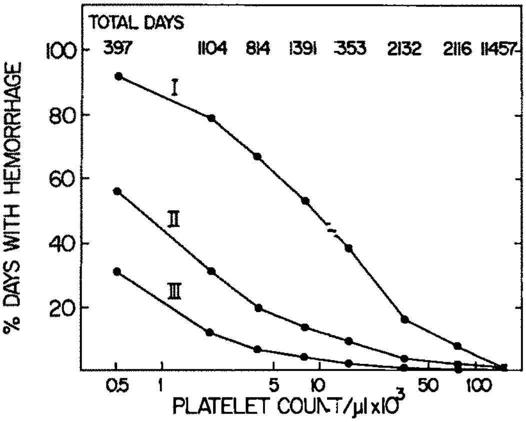


Figure 10 Relation between hemorrhage and platelet count. The percentage of days with hemorrhage for the 92 patients with acute leukemia. Figures across the top are the total number of patient days at each platelet count. Curve I shows data for all hemorrhagic manifestations. In curve II, skin hemorrhage and epistaxis are excluded. Curve III refers only to grossly visible hemorrhage (hematuria, melena, hematemesis). (From Ref. 63.)

quately resolved. The original studies of Freeman & Buckley (71) suggested that fever and thrombocytopenia were independent variables that were both associated with bleeding in patients with platelet counts of less than 60,000/µl. At higher platelet levels fever did not contribute to bleeding. Fever was positively correlated with elevation of serum polysaccharides derived from bacterial membranes. In most of their patients fever was a manifestation of infection and they concluded that the mechanism for fever increasing the bleeding risk could be related either to (*a*) a decreased platelet lifespan and thus more pronounced thrombocytopenia, (*b*) serum polysaccharides or some other fever-associated factor inhibiting platelet function, or (*c*) infection-damaged vessel walls requiring more platelets than the number available. A subsequent study by Roy & Djerassi (26) indicated that bleeding was not more common in infected vs noninfected leukemics and this lack of correlation was attributed to administration of different antibiotics or infection with another spectrum of agents. However, all of these patients were on prophylactic platelet transfusions so an effect of the infection may have been obscured. The detrimental role is most likely related to the actual presence of infection with additional platelet consumption and associated platelet dysfunction rather than to a direct result of fever per se.

These studies indicate that it is impossible to exactly define the indications for platelet transfusions to prevent bleeding based on platelet count because platelet dysfunction is so common. The critical question is whether bleeding can be adequately controlled as it occurs or whether a patient must be treated prophylactically by some arbitrarily defined criteria. In addition, how does the platelet transfusion program selected affect the patient's long-term prognosis? Clearly, indiscriminate use of platelet transfusions may lead to inability to control bleeding. For example, 30 patients (6 with solid tumors, 24 leukemics) were given 393 tranfusions during their treatment course (72): 82% of the transfusions were given for prophylaxis at platelet counts < 20,000/μl and 15% for bleeding. Table 1 shows that with increasing numbers of prior platelet transfusions the ability to control bleeding was markedly compromised. Only two studies have attempted to address the question of prophylactic vs therapeutic transfusions in a controlled fashion. Both gave platelet transfusions to a prophylactic group for platelet counts of less than 20,000/μl and the second group was transfused only for bleeding that involved more than skin, mucous membrane, or epistaxis (specific indications group) (73). Table 2 shows results from one of the studies indicating that there was no difference in the number of red cell transfusions, remissions or bleeding deaths between the two groups. However, the prophylactic group required twice the number of platelet transfusions to accomplish the same result as the therapeutically treated patients. In another investigation of 76 children with acute leukemia, the prophylactic group again required twice as many platelet transfusions (74). Days with active bleeding were significantly less in the prophylactic group for the first four months of study but were equal in the two groups by the end of the study. Several children in the prophylactic group did not have their bleeding controlled by transfusion in the last months of their illness conceivably because they had a higher incidence of alloimmunization than the therapeutic group. Survival was not different in the two groups, which indicates that the children transfused only for bleeding did not die before a transfusion was effective. These studies emphasize the concern that prophylactic trans-

Table 1 The effect of previous transfusions on the ability of platelet transfusions to control hemorrhage[a]

Number of prior platelet transfusions	Continued hemorrhage after transfusion		Total
	No	Yes	
<3	8 (62%)	5 (38%)	13
3–9	8 (42%)	11 (58%)	19
10–19	5 (56%)	4 (44%)	9
20–29	1 (8%)	11 (82%)	12
>30	0 —	2 (100%)	2
Totals	22 (40%)	33 (60%)	55

[a] From Ref. 72.

Table 2 Results of prophylactic vs therapeutic platelet transfusions[a]

	Patient group	
	Prophylactic[b]	Specific indications[c]
Age (yr) (mean ± S.E.)	45 ± 3	44 ± 5
Platelet packs per course (mean ± S.E.)	31[illegible] ± 5.9	16.1 ± 3.4
Red cell packs per course	6[illegible] ± 1.0	7.3 ± 1.0
Complete remissions	[illegible]/17	6/12
Complete and partial remissions	1[illegible]/17	10/12
All deaths within 1 month/course	3/22	2/17
Bleeding deaths within 1 month/course	[illegible]/22	0/17

[a] From Ref. 73.

[b] Platelet transfusions given to acute non-lymphoblastic leukemic patients with platelet counts of <20,000/μl.

[c] Platelet transfusions given to acute non-lymphoblastic leukemic patients only when bleeding involved more than skin, mucous membranes, or epistaxis.

fusion programs not only are more costly because of the increased number of transfusions but also are not cost effective, since no better results are achieved than with therapeutic transfusions. In addition, patients still requiring platelet support late in their illness may not be able to have adequate hemostasis because of alloimmunization.

Methods of Preventing Alloimmunization to Platelet Transfusions

DONOR SELECTION No systematic study of donor selection with long-term follow-up of sensitization rates in patients has been reported. However, it has been clearly established that HLA antigens are present on platelets and that when a patient becomes refractory to random donor platelet transfusions, HLA typing may be useful in selecting compatible platelet donors (75, 76). Whether initial selection of HLA compatible donors would prevent sensitization or at least extend the time before sensitization is an unanswered question. In order to address this problem we used an animal model that circumvented the clinical variables of underlying disease, immunosuppressive medications, infections, or other blood product transfusions and their potential effects on alloimmunization (77). Normal nonimmunosuppressed dogs were allocated to receive platelets from (*a*) a single random donor dog, (*b*) a pool of [illegible] random donors, (*c*) a DLA mismatched sibling, (*d*) a DLA matched sibling, or (*e*) the platelet pool until sensitization followed by a transfusion from a DLA matched sibling. ^{51}Cr-labeled donor platelets were transfused weekly into the recipient for 26 weeks or until sensitization, defined as zero labeled platelets circulating 24 hours post-infusion. Comparison of the percentage of sensitized dogs and the time to sensitization yielded the results shown in Table 3.

Table 3 Sensitization to selected donor platelets in a dog model[a]

Platelet donor	Recipients sensitized/transfused	Average number of transfusions until sensitized/(range)
Single random dog	13/14 (93%)	2.7/(2–7)
Pool of 6 random dogs (pool)	17/21 (81%)	7.4/(2–16)
DLA mismatched sibling	4/12 (33%)	8.3/(2–15)
DLA matched sibling	3/11 (27%)	19.0/(15–25)
DLA matched sibling after pool	5/15 (33%)	1

[a] From Ref. 77.

The following conclusions were drawn: (*a*) a platelet pool allows longer support than a single random donor (7.4 vs 2.7 weeks), (*b*) some random donor-recipient combinations do not immunize (7–19%), (*c*) platelet antigens responsible for alloimmunization are inherited but are not necessarily defined by the DLA locus, (*d*) in one third of the dogs, prior random transfusions sensitized the recipient to a DLA matched sibling, and (*e*) exclusive transfusion from a matched sibling provides the longest period of platelet support. It is of interest that approximately the same percentage of dogs became sensitized to a DLA matched littermate whether platelet transfusions had only been from the littermate (27%) or had been preceded by pooled random transfusions (33%). This suggests that prior random platelet transfusions may select out the DLA matched sibling pairs who will become immunized, but continued support from a truly compatible donor will not be compromised.

The data indicate that we cannot routinely predict compatible platelet donors on the basis of histocompatibility typing. This conclusion has been confirmed in patient studies (18, 78–80). Thus, the best use of resources would be to provide pooled random donor platelets until sensitization occurs, followed by support from a compatible donor.

REMOVAL OF LEUKOCYTES In a very interesting study, Herzig et al (81) observed that an incompatible platelet transfusion response could often be made compatible when 96% of the contaminating white blood cells were removed by a second centrifugation of the platelet concentrate at 178 g for 3 minutes. This procedure also removed 20% of the platelets. The compatibility improved regardless of the HLA matching grade of the donor with the recipient (Figure 11). In that study the median duration of responsiveness to the leukocyte-poor concentrates had not been reached at the time of writing but was in excess of 6 months (Figure 12), and there was no statistical difference between HLA match grades and duration of respon-

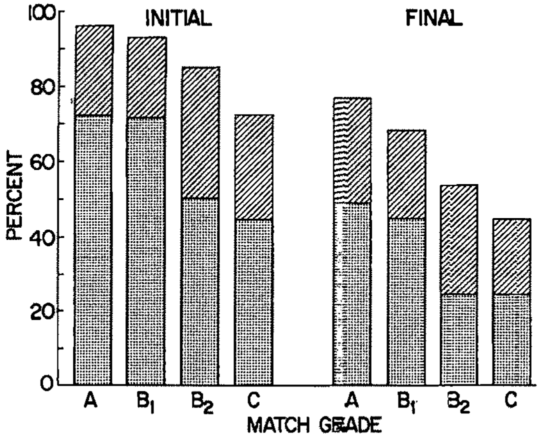


Figure 11 Relationship of transfused responses to processing HL-A selected platelet concentrates. The dotted portion of each bar represents the percentage of the 17 alloimmunized, thrombocytopenic patients responding to routine platelet concentrates from HLA selected donors (142 transfusions). The cross-hatched portion of the bars represents the additional percent of the group responding following transfusion of leukocyte-poor platelet concentrates prepared by recentrifuging the platelet concentrates at 178 g for 3 minutes. The total height of the bar represents the sum of the standard and leukocyte-poor concentrates and therefore the best possible response using both techniques. The initial response for the first transfusion of the selected HLA donor into the recipient is compared to the results following continued transfusions from the same donor. A match = donor and recipient have identical HL-A phenotype (or genotype). B match = all donor HL-A antigens are present in the recipient phenotype, but the donor lacks one (B-1) or two (B-2) of the recipient antigens. C match = the donor possesses one or more HL-A antigen(s) not found in the recipient. (From Ref. 81.)

siveness. This finding was in contrast to the leukocyte-rich concentrates in which duration of responses to HLA-A and B-1 matches was 4.5 months and to the lesser B-2 and C matches 1.5 months. These differences were statistically significant, $P < 0.01$. Thus, there appeared to be a very long duration of responsiveness to leuko-poor platelet concentrates which was not related to HLA matching grade. Furthermore, transfusion reactions of urticaria, fever, and/or chills were eliminated in 87% of the recipients of the leuko-poor concentrates. In addition, recipients of leukocyte-rich incompatible platelet concentrates often had post-transfusion granulocytopenia with subsequent infection (82).

In a subsequent study Eernisse & Brand (83) provided leuko-poor platelets (400 g for 7 minutes with 25% reduction in platelets) along with leukocyte-free red cells prepared using nylon wool columns to a group of thrombocytopenic patients. Sensitization rates in these patients were compared to a retrospective control group of patients given routinely prepared

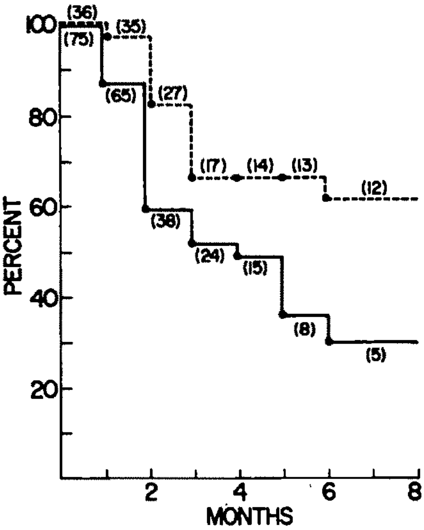


Figure 12 Duration of compatible responses to platelet concentrate transfusions. The solid line represents the duration using routine concentrates. The broken line represents the duration obtained with leukocyte-poor concentrates. The numbers in parentheses are the number of donor-recipient pairs at risk at each time interval. (From Ref. 81.)

platelet concentrates and buffy coat poor or leuko-pack filtered whole blood. In the control group 26 of 28 patients became immunized at 5 weeks (Table 4) while in the leuko-poor group only 16/68 became immunized and 52 were still not immunized after a median interval of 11 weeks.

Since publication of Herzig's study we have not observed a single patient in whom the additional centrifugation of the platelet concentrate reversed an incompatibility nor have we observed the development of post-transfusion granulocytopenia and infection. We have seen three patients in whom platelets were compatible as judged by increment and survival but the transfusion was associated with chills and fever. Using Herzig's additional centrifugation eliminated the transfusion reaction, but the platelet response was not further improved. The explanation for these discrepant data may be related to differences in the number of white cells in our platelet concentrates compared to that found in Eernisse's and Herzig's (Table 5) (84). This very important issue needs further clarification not only because sensitization may be delayed with leuko-poor platelets but also because it may explain some of the discrepancies in the importance of HLA typing to platelet compatibility. Herzig's data indicate that leuko-poor platelet concentrates are less dependent on HLA compatibility for a transfusion response (Figure 11) (81).

Table 4 Antigenicity of two platelet regimens (From Ref. 83)

	No. of patients	Transfusions Blood (units) ave./(range)	 Platelets (units) ave./(range)	Refractory	Transfusion duration (weeks) median/(range)
Unmodified	26	8 (0–20)	12 (0–67)	Yes	5 (3–15)
Platelet and blood transfusions	2	11 (8–14)	11 ([illegible]–16)	No	8 (6–10)
Leuko-poor	16	13 (2–43)	27 (0–90)	Yes	6 (2–53)
Platelet and blood transfusions	52	14 (4–46)	40 (12–110)	No	11 (6–32)

Table 5 Leukocyte contamination of routinely prepared vs recentrifuged platelet concentrate

	Platelets × 10^{11}	Leukocytes × 10^{8}
Herzig (81)		
routine	0.88	8.36
recentrifuged	0.66	0.29
Eernisse (83)		
routine	0.83	4.05[a]
recentrifuged	0.63	<0.01
Perkins (84)[b]		
routine	0.81	1.4

[a] Leukocyte contamination between 10–20% of the number present in whole blood (Brand, personal communication). Since recentrifugation of platelet concentrates in our laboratory does not improve the platelet increment in alloimmunized recipients, it suggests the critical level of wbc contamination is between 1.4 and 4.1 × 10^8 leukocytes per platelet concentrate. It also indicates that some routine methods of platelet preparation may maintain compatibility and eliminate the need for further recentrifugation of the platelet concentrates with the associated loss of platelets.

[b] Perkins-prepared platelet concentrates as described by Slichter & Harker, (13).

DISEASE SUPPRESSION OF ALLOIMMUNIZATION Based on a study by Freireich et al it has been generally accepted that leukemic patients rarely become immunized compared to patients with aplastic anemia (8). However, in Freireich's study, the recipients usually received platelets from a single (parent) donor and thus were less likely to become immunized. Furthermore, the 28 patients tested received an average of only 7 transfusions with a range of 1 to 19, which is often not enough to produce alloimmunization. Other evidence supporting disease-related differences in immunization was a study of Rh immunization in Rh negative leukemics receiving incompatible Rh positive platelet transfusions. Of 102 patients receiving platelets containing contaminating red cells, only 7.8% developed Rh antibodies compared to a 60–80% incidence in a normal control group receiving similar quantities of Rh positive red cells in a planned immunization program (85). However, since many of the leukemic patients were also receiving immunosuppressive drugs at the time of study, disease vs drug effects could not be distinguished. Other studies (86, 87) have shown ability to immunize as measured by a high incidence of either lymphocytotoxic antibodies or increasing titers of anti-A or -B in patients with leukemia. The antibody changes were directly related to the number of prior transfusions. These findings substantiate the original studies of Shulman (62) in which the number of red cell transfusions received was directly correlated with the formation of complement-fixing platelet antibodies. There is probably little basic difference between the immunologic responsiveness to transfusions between patients with leukemia and aplasia, particularly when it is considered that (*a*) leukemics often receive immunosuppressive drugs as part of their treatment program, and (*b*) antibody production should be a function not only of the number and type of transfusions but also of the transfusion schedule, which may vary significantly between leukemic and aplastic patients.

EFFECTS OF SPLENECTOMY OR DRUGS Several studies have analyzed the role of the spleen in impairing platelet transfusion response in alloimmunized patients. Flatow & Freireich (88) described three alloimmunized patients in whom splenectomy was associated with marked improvements in responses to random donor platelets. Conversely, Grumet & Yankee (47) found no effect of either splenectomy or steroids on refractory donors but did find improved responses to platelets from compatible donors, as would be expected with removal of the splenic storage pool. The explanation for the discrepancies in response to splenectomy between these studies is conceivably related to the level of circulating antiplatelet antibody. At least in patients with ITP, splenectomy is helpful if the patient has a low-titered antibody, but higher titers are associated with removal of the antibody-damaged cells in the liver rather than the spleen (89).

It has not been established that any drug given after alloimmunization can reverse the process. However, our preliminary studies in an animal model (Table 6) may be interpreted to indicate that some non-marrow toxic drugs when given before and during transfusions can delay or prevent sensitization.

Management of Refractory Alloimmumized Patients

SELECTION OF COMPATIBLE PLATELET DONORS The first extensive work on platelet alloantigens was done by Shulman et al (37) who described PLA^{-1}, a platelet specific antigen with a frequency of about 97% in the population tested. Van der Weerdt (90) extended these studies and de-

Table 6 Drug prevention of platelet sensitization in a baboon model

Drug administered[a]	Donor-recipient pairs (no.)	Time for sensitization (weeks)
None	6	2, 2, 3, 3, 4, 8
Antithymocyte serum (ATS)	3	3, >16, >38
Prednisone	2	16, >40
ATS + Prednisone	2	20, >28

[a] Drugs were given for one week before and daily during weekly platelet transfusions from a single unrelated donor animal.

scribed an additional platelet specific antigen system, KO. Subsequently, a PLGRLY^{B1} antigen was identified on platelets, granulocytes, and lymphocytes (37); this antigen was later found to be the same as the histocompatibility antigen now called HLA-A2. Colombani (91) and Svejgaard et al (92) established that the HLA antigens carried on an individual's lymphocytes are also expressed on the platelets. Although the HLA-A and -B antigens are present on platelets, some of them may have variable expression (93–95). The HLA-C, -D and Dr antigens have not been demonstrated on platelets (96, 97).

Because the shared platelet and lymphocyte HLA-A and -B antigens represent the major identifiable alloantigens on platelets, it was logical for Yankee et al to use HLA lymphocyte typing to select compatible platelet donors for alloimmunized thrombocytopenic patients. This method of donor selection was found to be successful both within the family (75) and in the random population (76). These studies have subsequently been confirmed by others (98–101).

However, the practicality of finding HLA identical (A) matches from within the general population is difficult because of the highly polymorphic nature of the human HLA genetic system. More than 50 antigens have been defined at the HLA-A and -B loci. Therefore, access to more than 50,000 HLA typed donors is necessary to assure a 50% probability of finding 10 perfectly matched donors for 100 potential recipients (102). As an approach to this problem, Duquesnoy et al (103) investigated the value of selecting donors with HLA antigens which were either the same as the patient's or were known to be crossreactive with the patient's antigens. They found that transfusions from such donors were useful when there was at least a BX match [i.e. the donor's antigens were the same as (B match) or at least crossreactive with the patient's (BX match)]. Of even more potential interest was the finding that some recipients responded very well to platelets from donors with major HLA differences, i.e. Grade C and D matches. In a Grade C match, the donor has one or two HLA antigens not present in the recipient; in a grade D match, three or four donor antigens differ from those of the recipient.

It was subsequently found by Duquesnoy et al (104) that those patients who responded well to C and D matched donor platelets often lacked the HLA-A2 antigen, while those who required A or B matched donors had HLA-A2. Although these studies suggest that donor selection within the HLA system permits some flexibility, the problem still remains that for any given donor-recipient combination one cannot necessarily predict the transfusion response. In fact, even a perfect HLA match may not always be indicative of success, because platelet alloimmunization to a matched sibling donor has been documented (78), and repeated transfusions from fully compatible unrelated A matched random donors is associated with unresponsiveness in up to a third of the recipients (79, 80, 105). These observations are important both from the standpoint of the patient who may continue to bleed if the transfusion is incompatible, and also the donor who may have had time-consuming, needless procedures.

Such findings have generated the need for studies to determine if there are other ways to select compatible donors. For example, Brand et al (79), using a combination of lymphocytotoxicity and fluorescent platelet antibody crossmatch techniques, reported a 93% success rate in predicting platelet transfusion compatibility. Similar results were obtained by Wu and colleagues (106–108), who used platelet aggregometry for compatibility testing. The latter technique has the disadvantage of requiring that fresh platelets from all potential donors be available at the time of testing. Numerous other techniques for platelet antibody detection are available, but their ability to select compatible platelet donors has not been specifically assessed (109–112). Although lymphocytotoxicity testing has been used extensively for this purpose, recent data suggest it has limited usefulness (107, 113).

Our own investigations (114) have included both platelet crossmatching tests and histocompatibility typing as methods of selecting donors for alloimmunized thrombocytopenic patients. These studies showed that a compatible family donor could be found for 19 of 36 alloimmunized recipients (53%) (Table 7). Selection of nonrelated compatible donors on the basis of HLA typing was possible but usually required HLA-A or -B grade matches. Two thirds of HLA-A or -B matched donors were successful. Platelet crossmatch tests (serotonin and adenine release and a radioimmunoassay for IgG on the platelet surface) could select compatible donors 84% of the time. Major emphasis needs to be placed on determining the relative expression and immunogenicity of the HLA antigens on platelets so that these can be matched as needed. In addition, standardization and simplification of platelet crossmatch tests to identify platelet specific antigens is required.

Table 7 Family platelet donors[a]

HLA matching grade of donor to recipient	Compatible[c]	Incompatible[c]	Recipients transfused (no.)
A match[b]	9	2	11
B match	1	2	2
C match	10	24	22
D match	0	6	4
Total	20	34	31

[a] From Ref. 18.

[b] These donors were all HLA-A, -B, and -D identical siblings.

[c] Results are for the first platelet transfusion given from family members (HLA match grades A through D) to 31 thrombocytopenic recipients who were alloimmunized to random donor platelets. B match donors had no HLA antigens in excess of the recipient, C match donors had one HLA antigen incompatible with the recipient, and D match donors were incompatible for 2 or more HLA antigens. An additional five patients in the study had no available family donors.

Depending on the facilities available to the clinician the best approach to management of the refractory patient may simply be to try available family members (parents, siblings or children) since about half the time a compatible donor can be found. Family members, however, should not be used if there is a possibility of a subsequent marrow transplant. Family member transfusions may sensitize against minor histocompatibility antigens, thus reducing the chances of a successful transplant (115, 116). If a compatible family donor cannot be found or transplantation is a consideration, the next step should be to utilize a random donor file of HLA typed community members willing to have a plateletpheresis procedure. It is in this situation that a combination of close HLA matching (A or B donors) along with platelet crossmatch tests between donor and recipient may yield a high percentage of compatible donors. If an HLA typed donor file is not available but platelet crossmatch tests are, then screening random donors may ultimately be successful in identifying a matched donor.

CONTINUOUS TRANSFUSION OF INCOMPATIBLE PLATELETS One of the major controversies in platelet transfusion therapy is whether the patient benefits from the infusion of viable platelets when there is no increment in the platelet count. Clearly, if a platelet transfusion results in a further decrement in the patient's own platelet count, conceivably as the result of an innocent bystander phenomenon, this is more detrimental than no transfusion.

Whether an incremental response is needed may depend on whether the clinical situation requires platelet plug formation to control bleeding or, alternatively, if only platelet-endothelial cell interaction is needed (117).

The bleeding time is the best measure of platelet plug formation and reflects the number of platelets required to prevent bleeding from disrupted capillaries. There is a direct inverse relationship between the circulating platelet count and the bleeding time (Figure 3), which becomes unmeasurable at platelet counts below 10,000–20,000/μl (51). Data from post-surgical patients suggest that if larger vessels are damaged, platelet counts up to 100,000/μl may be required to control bleeding (32). However, in stable aplastic patients not on any medication, stool blood loss (as a measure of bleeding through intact vessels) does not consistently occur until the platelet count is less than 5000/μl (Figure 9). Thus, in the refractory patient, continued incompatible transfusions may be helpful only if oozing through intact capillaries is the only defect being managed.

The fact that platelets may have some endothelial supportive function comes from several lines of evidence: (*a*) platelet survival is less in patients with thrombocytopenia than in normal controls (18–21, 27); (*b*) often there is a better increment with the second platelet transfusion in a severely thrombocytopenic patient than with the first: conceivably the first transfusion provides platelets that interact with the endothelium rather than circulate (9, 72, 118); and (*c*) the control of bleeding often extends beyond the correction of either the bleeding time or the platelet count (5, 19, 119, 120). Furthermore, in the irradiated dog model, small amounts of transfused platelets, which neither increased the platelet count nor corrected the bleeding time, prevented leakage of red cells through the vessels into the lymphatic system (26).

This discussion is *not* to be interpreted as a recommendation for routinely transfusing platelets into patients who do not have an increment in the platelet count or improve their template bleeding times. Rather it is a careful examination of the available data which suggests that in some patients with very low platelet counts (<5000 platelets/μl), capillary bleeding, and for whom no compatible donors are available, a clinical trial of frequent small platelet transfusions to see if bleeding is controlled may be justified (54).

Literature Cited

1. Duke, W. W. 1910. The relation of blood platelets to hemorrhagic disease: Description of a method for determining the bleeding time and coagulation time and report of three cases of hemorrhagic disease relieved by transfusion. *J. Am. Med. Assoc.* 55:1185
2. Dillard, G. H. L., Brecher, G., Cronkite, E. P. 1951. Separation, concentration, and transfusion of platelets. *Proc. Soc. Exp. Biol. Med.* 78:796
3. Stefanini, M., Chatterjea, J. B., Dameshek, W. 1952. Studies on platelets II. The effect of transfusion of platelet-rich polycythemic blood on the platelets and hemostatic function in "idiopathic" and "secondary" thrombocytopenic purpura. *Blood* 7:53
4. Stefanini, M., Dameshek, W. 1953. Collection, preservation and transfusion of platelets. *N. Engl. J. Med.* 248:797
5. Hirsch, E. O., Gardner, F. H. 1952. The transfusion of human blood platelets. *J. Lab. Clin. Med.* 39:556
6. Levin, R. H. 1964. Handbook for platelet transfusion therapy. *Special Publication,* National Cancer Institute, June
7. Han, T., Stutzman, L., Cohen, E., Kim, U. 1966. Effect of platelet transfusion on hemorrhage in patients with acute leukemia. *Cancer* 19:1937–42
8. Freireich, E. J., Kliman, A., Gaydos, L. A., Mantel, N., Frei, E. 1963. Response to repeated platelet transfusion from the same donor. *Ann. Intern. Med.* 59:277
9. Alvarado, J., Djerassi, I., Farber, S. 1965. Transfusion of fresh concentrated platelets to children with acute leukemia. *J. Pediatr.* 67:13–22
10. Higby, D. J., Cohen, E., Holland, J. F., Sinks, L. 1974. The prophylactic treatment of thrombocytopenic leukemic patients with platelets: a double blind study. *Transfusion* 14:440–46
11. Chappell, W. S. 1966. Platelet concentrates from acidified plasma: A method of preparation without the use of additives. *Transfusion* 6:308
12. Mourad, N. 1968. A simple method for obtaining platelet concentrates free of aggregates. *Transfusion* 8:48
13. Slichter, S. J., Harker, L. A. 1976. Preparation and storage of platelet concentrates. I. Factors influencing the harvest of viable platelets from whole blood. *Br. J. Haematol.* 34:395
14. Baldini, M., Costea, N., Dameshek, W. 1969. The viability of stored human platelets. *Blood* 26:1669
15. Kissmeyer-Nielsen, F., Gedergaard, J. 1961. Platelets in blood stored in untreated and silconized glass bottles and plastic bags. I. Studies in vitro, and II. Survival studies. *J. Clin. Pathol.* 14:626, 630
16. Levin, R. H., Freireich, E. J., Chappell, W. 1964. Effect of storage up to 48 hours on response to transfusion of platelet rich plasma. *Transfusion* 4:251
17. Slichter, S. J. 1978. Efficacy of platelets collected by semi-continuous flow centrifugation (Haemonetics Model 30). *Br. J. Haematol.* 38:131–40
18. Slichter, S. J., Harker, L. A. 1978. Thrombocytopenia: Mechanisms and management of defects in platelet production. *Clin. Haematol.* 7:523–39
19. McGovern, J. J. 1957. Platelet transfusions in pediatrics. *N. Engl. J. Med.* 256:922
20. Gardner, F. H., Howell, D., Hirsch, E. O. 1954. Platelet transfusions utilizing plastic equipment. *J. Lab. Clin. Med.* 43:196
21. Mustard, J. F., Murphy, E. A., Robinson, G. A., Rowsell, H. C., Ozge, A. H., Crookston, J. H. 1964. Blood platelet survival. *Thromb. Diathes. Haemorrh. (Stuttgart)* Suppl. 13:245
22. Aursnes, I. 1974. Blood platelet production and red cell leakage to lymph during thrombocytopenia. *Scand. J. Haematol.* 13:184
23. Gimbrone, M. A., Aster, R. H., Cotran, S. 1969. Preservation of vascular integrity in organs perfused in vitro with a platelet-rich medium. *Nature* 222:33
24. Wojcik, J. D., van Horn, D. L., Webber, A. J., Johnson, S. A. 1969. Mechanism whereby platelets support the endothelium. *Transfusion* 9:324
25. Tranzer, J. P., Baumgartner, H. R. 1967. Filling gaps in the vascular endothelium with blood platelets. *Nature* 216:1126
26. Roy, A. J., Djerassi, I. 1972. Effects of platelet transfusions: Plug formation and maintenance of vascular integrity. *Proc. Soc. Exp. Biol. Med.* 139:137
27. Mustard, J. F., Rowsell, H. C., Murphy, E. A. 1966. Platelet economy (platelet survival and turnover). *Br. J. Haematol.* 12:1
28. Harker, L. A. 1971. The role of the spleen in thrombokinetics. *J. Lab. Clin. Med.* 77:247
29. Harker, L. A. 1970. Thrombokinetics in idiopathic thrombocytopenic purpura. *Br. J. Haematol.* 19:95
30. Harker, L. A., Slichter, S. J. 1972. Platelet and fibrinogen consumption in man. *N. Engl. J. Med.* 287:999–1005

31. Simon, T. L., Curreri, P. W., Harker, L. A. 1977. Kinetic characterization of hemostasis in thermal injury. *J. Lab. Clin. Med.* 89:702
32. Counts, R. B., Haisch, C., Simon, T. L., Maxwell, N. G., Heimbach, D. M., Carrico, C. J. 1979. Hemostasis in massively transfused trauma patients. *Ann. Surg.* 190:91–99
33. Harker, L. A., Branson, H., Slichter, S. J. Unpublished observations.
34. Slichter, S. J., Harker, L. A. 1974. Hemostasis in malignancy. *Ann. NY Acad. Sci.* 230:252–61
35. Shulman, N. R. 1958. Immunoreactions involving platelets. I. A steric and kinetic model for formation of a complex from a human antibody, quinidine as a haptene and platelets; and for fixation of complement by the complex. *J. Exp. Med.* 107:665
36. Gynn, T. N., Messmore, H. L., Friedman, I. A. 1972. Drug-induced thrombocytopenia. *Med. Clin. North Am.* 56:65–79
37. Shulman, N. R. Marder, V. J., Hiller, M. C., Collier, E. M. 1964. Platelet and leukocyte isoantigens and their antibodies: serologic, physiologic and clinical studies. *Prog. Hematol.* 4:222
38. Cimo, P. L., Aster, R. H. 1972. Post-transfusion purpura. *N. Engl. J. Med.* 287:290
39. Gardner, F. H. 1974. Use of platelet transfusions. *Br. J. Haematol.* 27:537
40. Murphy, S. 1972. Intrinsic platelet defects in hereditary thrombocytopenia. *Ann. NY Acad. Sci.* 20:421
41. Jackson, D. P., Sorensen, D. K., Cronkite, E. P., Bond, V. P., Fliedner, T. M. 1959. Effectiveness of transfusions of fresh and lyophilzed platelets in controlling bleeding due to thrombocytopenia. *J. Clin. Invest.* 38:1689–96
42. Hjort, P. F., Perman, V., Cronkite, E. P. 1959. Fresh, disintegrated platelets in radiation thrombocytopenia. *Proc. Soc. Exp. Biol. Med.* 102:31–34
43. Fliedner, T. M., Sorensen, D. K., Bond, V. P., Cronkite, E. P., Jackson, D. P., Adamik, E. 1958. Comparative effectiveness of fresh and lyophilized platelets in controlling irradiation hemorrhage in the rat. *Proc. Soc. Exp. Biol. Med.* 99:731–33
44. Firkin, B. G., Arimura, G., Harrington, W. J. 1960. A method for evaluating the hemostatic effect of various agents in thrombocytopenic rats and mice. *Blood* 15:388–93
45. Raccuglia, G., Bethell, F. H. 1960. Platelet transfusions and administration of platelet derivatives in man. *Am. J. Clin. Pathol.* 34:505–12
46. Cavins, J. A., Farber, S., Roy, A. J. 1968. Transfusion of fresh platelet concentrates to adult patients with thrombocytopenia. *Transfusion* 8:24–26
47. Grumet, F. G., Yankee, R. A. 1970. Long-term platelet support of patients with aplastic anemia. *Ann. Intern. Med.* 73:1–7
48. Freireich, E. J. 1978. Transfusional therapy for cancer patients. *Curr. Prob. Cancer* 3:1–37
49. Cohen, P., Gardner, F. H. 1966. Thrombocytopenia as a laboratory sign and complication of gram-negative bacteremic infection. *Arch. Intern. Med.* 117:113
50. Slichter, S. J. 1978. Preservation of platelet viability and function during storage of concentrates. 10th Annual Red Cross Scientific Symposium. *The Blood Platelet in Transfusion,* pp. 83–100. New York: Liss
51. Harker, L. A., Slichter, S. J. 1972. The bleeding time as a screening test for evaluating platelet function. *N. Engl. J. Med.* 287:155–59
52. Murphy, S., Gardner, F. H. 1969. Platelet preservation. Effect of storage temperature on maintenance of platelet viability—deleterious effect of refrigerated storage. *N. Engl. J. Med.* 280:1094
53. Slichter, S. J., Harker, L. A. 1976. Preparation and storage of platelet concentrates. II. Storage variables influencing platelet viability and function. *Br. J. Haematol.* 34:403–19
54. Högman, C. F., Krijnen, H. W., Valeri, C. R., eds. 1974. *Platelet Preservation and Transfusion,* pp. 5–113. Uppsala, Sweden: International Society of Blood Transfusion
55. Murphy, S., Sayar, S. N., Gardner, F. H. 1970. Storage of platelet concentrates at 22°C. *Blood* 35:549
56. Becker, G. A., Tuccelli, M., Kunicki, T., et al. 1973. Studies of platelet concentrates stored at 22C and 4C. *Transfusion* 13:61
57. Filip, D. J., Aster, R. H. 1978. Relative hemostatic effectiveness of human platelets stored at 4°C and at 22°C. *J. Lab. Clin. Med.* 91:618
58. Handin, R. I., Valeri, C. R. 1971. Hemostatic effectiveness of platelets stored at 22°C. *N. Engl. J. Med.* 285:538
59. O'Brien, J. R. 1968. Effects of salicylates on human platelets. *Lancet* 1:779
60. Valeri, C. R. 1974. Hemostatic effectiveness of liquid-preserved and previ-

ously frozen human platelets. *N. Engl. J. Med.* 290:353

61. Kevy, S. V., Jacobson, M. S., Button, L. N. 1976. The hemostatic effectiveness of supernatant plasma of 4°C platelet concentrates. *Blood* 48:1004
62. Shulman, N. R. 1966. Immunological considerations attending platelet transfusion. *Transfusion* 6:39–49
63. Gaydos, L. A., Freireich, E. J., Mantel, N. 1962. The quantitative relation between platelet count and hemorrhage in patients with acute leukemia. *N. Engl. J. Med.* 266:905–9
64. Roy, A. J., Jaffe, N., Djerassi, I. 1973. Prophylactic platelet transfusions in children with acute leukemia: A dose response study. *Transfusion* 13:283–90
65. Van Der Weyden, M. B., Clancy, R. L., Howard, M. A., Firkin, B. G. 1972. Qualitative platelet defects with reduced life-span in acute leukemia. *Aust. N. Z. J. Med.* 4:339–45
66. Cardamon, J. M., Edson, R., McArthur, J., Jacob, H. 1972. Abnormalities of platelet function in the myeloproliferative disorders. *J. Am. Med. Assoc.* 221:270
67. McKay, D. G. 1965. *Disseminated Intravascular Coagulation.* New York: Harper & Row
68. Brown, C. H., Bradshaw, M. W., Natelson, E. A., Aflrey, C. P., Williams, T. W. 1976. Defective platelet function following the administration of penicillin compounds. *Blood* 47:949–56
69. Brown, C. H., Natelson, E. A., Bradshaw, M. W., Williams, T. W., Alfrey, C. P. 1974. The hemostatic defect produced by carbenicillin. *N. Engl. J. Med.* 291:265–70
70. Weiss, H. J., Aldeort, L. M., Kochwa, S. 1968. The effect of salicylates on the hemostatic properties of platelets in man. *J. Clin. Invest.* 47:2169–80
71. Freeman, G., Buckley, E. S. 1954. Serum polysaccharide and fever in thrombocytopenic bleeding in leukemia. *Blood* 9:586–95
72. van Eys, J., Thomas, D., Olivos, B. 1978. Platelet use in pediatric oncology: A review of 393 transfusions. *Transfusion* 18:169–73
73. Solomon, J., Bokefkamp, T., Fahey, J. L., Chillar, R. K., Beutler, E. 1978. Platelet prophylaxis in acute non-lymphocytic leukemia. *Lancet* 1:267
74. Murphy, S., Litwin, S., Koch, P., Remischovsky, J., Donaldson, M. H., Evans, A. E., Gardner, F. H. 1976. The indications for platelet transfusion in children with acute leukemia. *Clin. Res.* 24:379A
75. Yankee, R. A., Grumet, F. C., Rogentine, G. N. 1969. Platelet transfusion therapy. The selection of compatible platelet donors for refractory patients by lymphocyte HLA-A typing. *N. Engl. J. Med.* 281:1208
76. Yankee, R. A., Graff, K. S., Dowling, R. 1973. Selection of unrelated compatible platelet donors by lymphocyte HL-A matching. *N. Engl. J. Med.* 288:760
77. O'Donnell, M. R., Slichter, S. J. 1979. Platelet (Plt) alloimmunization—Correlation with donor source. *Am. Fed. Clin. Res.* 390 (Abstr.)
78. Bucher, U., deWeck, A., Spengler, H. 1973. Platelet transfusions: Shortened survival of HL-A-identical platelets and failure of in vitro detection of antiplatelet antibodies after multiple transfusions. *Vox Sang.* 25:187
79. Brand, A., van Leeuwen, A., Eernisse, J. G. 1978. Platelet transfusion therapy. Optimal donor selection with a combination of lymphocytoxicity and platelet fluorescence tests. *Blood* 51:781–88
80. Herzig, R. 1975. Histocompatible platelet transfusions: The NIH experience. *Single Donor Platelet Transfusions. Scientific, Legal, and Ethical Considerations,* 76-1059, p. 12. Washington, DC: NHEW
81. Herzig, R. H., Herzig, G. P., Bull, M. I., Decter, J. A., Lohrmann, H.-P., Stout, F. G., Yankee, R. A., Graw, R. G. 1975. Correction of pooled platelet transfusion responses with leukocyte-poor HL-A-matched platelet concentrates. *Blood* 46:743–50
82. Herzig, R. H., Poplack, D. G., Yankee, R. A. 1974. Prolonged granulocytopenia from incompatible platelet transfusions. *N. Engl. J. Med.* 290:1220–23
83. Eernisse, J. G., Brand, A. 1980. Postponement of sensitization against HLA antigens by administration of leucodepleted blood components. *J. Exp. Hematol.* In press
84. Perkins, H. A. Personal communication
85. Goldfinger, D., McGinniss, M. H. 1964. Rh-Incompatible platelet transfusions—risks and consequences of sensitizing immunosuppressed patients. *N. Engl. J. Med.* 284:942–44
86. Schiffer, C. A., Lichtenfeld, J. L., Wiernik, P. H., Mardiney, M. R., Joseph, J. M. 1976. Antibody response in patients with acute nonlymphocytic leukemia. *Cancer* 37:2177–82

87. Tejada, F., Bias, W. B., Santos, G. W., Zieve, P. D. 1973. Immunologic response of patients with acute leukemia to platelet transfusions. *Blood* 42: 405–12
88. Flatow, F. A., Freireich, E. J. 1966. Effect of splenectomy on the response to platelet transfusion in three patients with aplastic anemia. *N. Engl. J. Med.* 274:242–48
89. Aster, R. H., Jandl, J. A. 1964. Platelet sequestration in man. II. Immunological and clinical studies. *J. Clin. Invest.* 43:856–69
90. Van der Weerdt, C. M. 1965. *Platelet antigens and isoimmunization.* Thesis, Drukkenis Aemstalsted, Amsterdam
91. Colombani, J. 1971. Blood platelets in HL-A serology. 1971. *Transplant Proc.* 3:1078
92. Svejgaard, A., Kissmeyer-Nielsen, F., Thorsby, E. 1970. HL-A typing of platelets. In *Histocompatibility Testing 1970,* ed. P. I. Terasaki, p. 160. Copenhagen: Munksgaard
93. Liebert, M., Aster, R. H. 1977. Expression of HLA-B12 on platelets, on lymphocytes and in serum. A quantitative study. *Tissue Antigens* 9:199
94. Aster, R. H., Szatkowski, N., Liebert, M. 1977. Expression of HLA-B12, HLA-B8, w4, and w6 on platelets. *Transplant. Proc.* 9:1965
95. Duquesnoy, R. J., Testin, J., Aster, R. H. 1977. Variable expression of w4 and w6 on platelets. Possible relevance to platelet transfusion therapy of alloimmunized thrombocytopenic patients. *Transplant Proc.* 9:1829
96. Duquesnoy, R. J., Filip, D. J., Tomasulo, P. A. 1977. Role of HLA-C matching in histocompatible platelet transfusion therapy of alloimmunized thrombocytopenic patients. *Transplant. Proc.* 9:1827
97. van Rood, J. J., van Leeuwen, A., Keuning, J. J. 1975. The serological recognition of the human MLC determinants using a modified cytotoxicity technique. *Tissue Antigens* 5:73
98. Mital, K. K., Ruder, E. A., Green, D. 1976. Matching of histocompatibility (HL-A) antigens for platelet transfusion. *Blood* 47:31
99. Lohrmann, H., Bull, M. I., Decter, J. A. 1974. Platelet transfusions from HL-A compatible unrelated donors to alloimmunized patients. *Ann. Intern. Med.* 80:9
100. Radvany, R., Green, D., Rossi, E. C. 1977. Efficacy of matched platelet transfusions from unrelated donors. *Transplant. Proc.* 9:513
101. Thorsby, E., Helgeson, A., Gjemdal, T. 1972. Repeated platelet transfusions from HLA compatible unrelated donors. *Tissue Antigens* 2:397
102. Opelz, I., Mickey, M. R., Terasaki, P. 1974. Unrelated donors for bone marrow transplantation and transfusion support: Pool sizes required. *Transplant. Proc.* 4:405
103. Duquesnoy, R. J., Filip, D. J., Rodey, G. E. 1977. Successful transfusion of platelets "mismatched" for HLA antigens to alloimmunized thrombocytopenic patients. *Am. J. Hematol.* 2:219
104. Duquesnoy, R. J., Filip, D. H., Aster, R. H. 1977. Influence of HLA-A2 on the effectiveness of platelet transfusions in alloimmunized thrombocytopenic patients. *Blood* 50:407
105. Tosato, G., Applebaum, F. R., Deisseroth, A. B. 1978. HLA-matched platelet transfusion therapy of severe aplastic anemia. *Blood* 52:846–54
106. Wu, K. K., Hoak, J. C., Thompson, J. S. 1975. Use of platelet aggregometry in selection of compatible platelet donors. *N. Engl. J. Med.* 292:130
107. Wu, K. K., Hoak, J. C., Koepke, J. A. 1977. Selection of compatible platelet donors: A prospective evaluation of three cross-matching techniques. *Transfusion* 17:638
108. Wu, K. K., Thompson, J. S., Koepke, J. A. 1976. Heterogeneity of antibody response to human platelet transfusion. *J. Clin. Invest.* 58:432
109. Hirschman, J. R., Shulman, N. R. 1973. The use of platelet serotonin release as a sensitive method for detecting antiplatelet antibodies and a plasma antiplatelet factor in patients with idiopathic thrombocytopenic purpura. *Br. J. Haematol.* 24:793
110. Hirschman, R. J., Gralnick, H. R. 1974. A simplified platelet factor 3 (PF-3) assay for the rapid detection of platelet isoantibodies and an antiplatelet factor in ATP and SLE. *J. Lab. Clin. Med.* 84:292
111. Heinrich, D., Kessler, C., Mueller-Eckhardt, C. 1977. A comparative analysis of methods suitable for detection of platelet alloantibodies. *Vox Sang.* 32:325
112. Tate, D. Y., Sorenson, R. L., Gerrard, J. M. 1977. An immunoenzyme histochemical technique for the detection of platelet antibodies from the serum of patients with idiopathic (autoimmune)

thrombocytopenic purpura (ITP). *Br. J. Haematol.* 37:265

113. Herzig, R. H., Terasaki, P. I., Trapani, R. J. 1977. The relationship between donor-recipient lymphocytotoxicity and the transfusion response using HLA-matched platelet concentrates. *Transfusion* 17:657
114. Slichter, S. J. 1978. Selection of compatible platelet donors. In *Platelet Physiology and Transfusion,* pp. 83–92. Washington, DC: Am. Assoc. Blood Banks
115. Storb, R. S., Thomas, E. D., Weiden, P., et al. 1976. Aplastic anemia treated by allogenic bone marrow transplantation, report of 49 new cases from Seattle. *Blood* 48:817–41
116. Storb, R. S., Weiden, P., Deeg, H. J., et al. 1979. Rejection of marrow from DLA identical canine littermates given transfusions before grafting: Antigens involved are expressed on leukocytes, skin epithelial cells but not on platelets and red blood cells. *Blood* 54:477–84
117. Johnson, S. A., Van Hord, D. L., Pederson, H. J., Marr, J. 1966. The function of platelets: a review. *Transfusion* 6:3–17
118. Djerassi, I., Farber, S., Evans, A. E. 1963. Transfusion of fresh platelet concentrates to patients with secondary thrombocytopenia. *N. Engl. J. Med.* 268:221–26
119. Castaldi, P. A., Firkin, B. G. 1964. Platelet transfusion. *Med. J. Aust.* 1:861–65
120. Tobin, J. R., Friedman, I. A. 1960. Platelet transfusion from use of blood in plastic bags from routine storage. *J. Am. Med. Assoc.* 172:100–2

Ann. Rev. Med. 1980. 31:541–60
Copyright © 1980 by Annual Reviews Inc. All rights reserved

INFANT BOTULISM

❖7377

Stephen S. Arnon, M.D.

Infant Botulism Research Project, Infectious Disease Section,
California Department of Health Services, Berkeley, California 94704

INTRODUCTION AND DEFINITIONS

Infant botulism is the newly-recognized form of botulism that results when spores of *Clostridium botulinum* germinate and then multiply in the gut and produce botulinal toxin in vivo (1). Infant botulism should be distinguished from the two other better-known forms of botulism. Food-borne botulism is an intoxication that results when an improperly preserved food that contains preformed botulinal toxin is eaten. Wound botulism, a rarely reported entity, results when *C. botulinum* organisms infect and produce toxin in traumatized tissue.

After absorption and hematogenous distribution, botulinal toxin binds presynaptically at the neuromuscular junction and other peripheral cholinergic synapses, thereby preventing acetylcholine release. The clinical effect of this action is hypotonia and a descending, symmetric flaccid paralysis. For as yet unknown reasons, the intestinal tract of only some infants is susceptible to *C. botulinum* infection; most infants, older children, and adults regularly ingest *C. botulinum* spores without ill effects.

HISTORY

Infant botulism is not a new disease; it is only newly recognized. The first case of human infant botulism—proved by laboratory examination—is now known to have occurred in California in 1931, although it was misdiagnosed at the time (2). This case is of particular note because it illustrates how concepts that are well-established and widely accepted at a particular time may block perception of newer and more correct perspectives. In 1919 a Botulism Commission, consisting of Meyer, Dickson, and Geiger, was created because explosive outbreaks of food-borne botulism, then a poorly-

0066-4219/80/0401-0541$01.00

understood disease, threatened both the public's health and the canning industry (3). In three years the Commission established much fundamental knowledge about botulism as a food-borne disease (4). A decade later in 1931 each of the former commissioners became personally involved with the case of the first patient to have infant botulism: Dickson tested the feces that established the diagnosis, Meyer tested the remaining cans of evaporated milk, and Geiger, as Health Officer of San Francisco, received the attending physician's report of the case (2). Although no food eaten by the patient could be proved to contain preformed botulinal toxin, the concept of botulism as a food-borne disease apparently had been so securely established by the Commissioners that they did not seriously consider in 1931 the possibility that *C. botulinum* could colonize the infant gut and produce toxin. Forty years later, however, when reminded of the case by the patient's father, Meyer at once perceived this possibility and planned experiments to test the idea (2).

The characteristic clinical features of infant botulism had become evident to some well before the etiology of the disease was understood. In 1974 Grover, Peckham & Berman (5) described from Pennsylvania nine patients with a neurologic syndrome of undetermined etiology that from today's perspective was almost certainly infant botulism. The same idiopathic syndrome was recognized in southern California and in 1976 was reported by Ramseyer, Clay & Fishman (6) to have a characteristic electromyographic pattern. In 1977 Clay et al (7) connected their eight patients to infant botulism.

The first report of overt botulism in two infants was that by Pickett et al in 1976 (8). The source of the botulinal toxin for these two patients was unknown (8), but the possibility of its in vivo production was suggested (8, 9). The clinical diagnosis of botulism in these and other infants was unambiguously confirmed by the identification of *C. botulinum* toxin and organisms in the patients' feces (9). Evidence was also obtained that ingested spores of *C. botulinum* produced the toxin in the infants' intestinal tracts (1, 9).

ETIOLOGIC AGENT

Clostridium botulinum is a Gram-positive, spore-forming, obligate anaerobe whose natural habitat worldwide is the soil. Consequently, *C. botulinum* is as ubiquitous as dust, and its spores are commonly present on fresh fruits, vegetables, and other agricultural products. The members of this species are so diverse in their biochemical capabilities that they would not be classed as a single species were it not for the similar toxin molecule each produces (10). In general, each vegetative cell of *C. botulinum* produces just one of seven serologically distinguishable toxins, arbitrarily designated by the letters A–G. All the toxin molecules are pure proteins, weigh approximately 150,000 daltons, and act by the same mechanism. The different toxin types serve as convenient epidemiologic and clinical markers.

Botulinal toxin is the most potent poison known. By extrapolation from studies with other animal species, the lethal dose of botulinal toxin in the bloodstream of man is approximately 10^{-9} mg per kg of body weight (11). Because the physiologic function of botulinal toxin to the parent bacterium has not been established, several investigators have wondered why the capacity to produce toxin has been retained by *C. botulinum* during some three billion years of evolution. The answer may derive from the recognition of human infant botulism, a potentially fatal disease if untreated, and from the occurrence of botulism in other animals. If the animal host dies from botulinal toxin, then a large mass of decomposing tissue becomes immediately available to support the growth and multiplication of *C. botulinum.* When nutrients become limiting, the now enlarged population of *C. botulinum* may be induced to sporulate. As the corpse disintegrates, the spores may be disseminated by wind and by scavengers. In this manner its toxin would provide *C. botulinum* with a means of increasing its numbers and its territory.

ANIMAL MODELS OF PATHOGENESIS

Several animal models have been developed that provide, by analogy, some insight into the pathogenesis of infant botulism. It is historically noteworthy that the capacity of toxin-free *C. botulinum* spores to colonize the gut and cause fatal botulism was first demonstrated in adult guinea pigs by Orr over half a century ago (12). Orr had to use large numbers (10^9) of spores to produce intestinal infection, presumably because of differences between the adult and the infant gut.

More recently, two other animal models have been developed. Miyazaki & Sakaguchi (13) fed baby chickens toxin-free *C. botulinum* spores and produced fatal botulism. These investigators showed that in this species the cecum is the site of spore germination and toxin production, as ligation of the cecum at seven days of age rendered the animals immune to subsequent oral challenge with toxin-free spores.

Using infant mice as a model, Sugiyama & Mills (14) experimentally reproduced the limited age of susceptibility to infant botulism that had been observed clinically (1, 11). They found that *C. botulinum* spores, when injected intragastrically, would germinate in the infant mouse gut only if the animal were between 7 and 13 days of age. Interestingly, although toxin was identified in the lumen of the colon, the mice in which the spores had germinated remained well (14). Subsequent studies suggested that the intraluminal toxin was not being absorbed (15). When germ-free mice were studied, Moberg & Sugiyama (16) found that the minimum age of susceptibility decreased from seven to three days. Adult germ-free mice could be colonized by as few as 10 spores, whereas ordinary adult mice could not be colonized even by 10^5 spores. Also, when adult germ-free mice were placed in the ordinary animal holding room for three days, they then became resistant to challenge with 10^5 spores (16). These experimental studies suggest that some organisms in the intestinal microflora contribute importantly to the enteric environment that ordinarily keeps *C. botulinum* spores from multiplying in the gut lumen.

EPIDEMIOLOGY

As of January 1980, infant botulism had been recognized in North America, Europe (England) (17), and Australia (18). Since *C. botulinum* spores have been identified on all continents, infant botulism is almost certainly global in distribution. All cases of infant botulism recognized to date have been caused by either type A or type B botulinal toxin.

Because infant botulism was so recently recognized as a distinct entity, awareness of its existence remains limited, and many presently occurring cases remain incorrectly diagnosed, as in past years (2, 19). In the United States, cases have thus far been recognized in 23 of the 50 states, located variously on the Pacific and Atlantic coasts, the Canadian and Mexican borders and in between. A national seasonality, if any, is not yet apparent from 139 laboratory-confirmed cases in the United States known as of January 1980. In general, the distribution of United States cases by toxin type has followed the established distribution of *C. botulinum* in American soil (20, 21); cases caused by type B toxin predominate east of the Mississippi River, while cases caused by type A toxin predominate west of it.

More cases have been recognized in California than in any other state, probably because of its special surveillance. However, when annual incidence is adjusted for the birth rate, Utah's incidence of hospitalized cases of infant botulism (about 3/10,000 live births) is the highest in the nation (22) and is approximately three times that of California. Because awareness of infant botulism remains limited, the actual incidence of infant botulism is at present unknown. The Center for Disease Control estimates that as many as 250 cases needing hospital care may be occurring in the United States annually (23), and a similar number may be dying unexpectedly at home before signs of illness become evident (11).

The most striking epidemiologic feature of infant botulism is its restricted age range. Of all recognized cases, 98% have occurred in patients between one and six months of age, and the oldest known case occurred in a nine-month-old child. All major racial and ethnic groups have been affected and both sexes equally so.

Investigation of potential host and environmental risk factors for infant botulism by means of a longitudinal case-control study was begun by the California State Health Department immediately after it recognized in 1976 that infant botulism was an infectious disease. In the California studies, a case of infant botulism was defined as (*a*) an illness (fatal or nonfatal) consistent with the effects of botulinal toxin and (*b*) the isolation of *C. botulinum* from fecal or autopsy specimens.

Because of the repeated identification of *C. botulinum* spores (but not toxin) in honey fed to infants who subsequently developed infant botulism (24, 25), and because a major honey producer chose to advise the public of this fact in July 1978 (26), an interim analysis of the California data was done at that time (27). Examination of the food histories of hospitalized cases and their controls revealed that honey exposure constituted a significant risk factor for type B infant botulism, as 44% of cases, vs only 9% of their controls, had been fed honey ($P = 0.005$). Furthermore, in every instance in which *C. botulinum* spores were isolated from the honey actually fed the patient, the toxin type of the patient's illness (A or B) matched the toxin type of the *C. botulinum* spore found in the honey (27). Honey was the only one of several hundred food and drug items tested from which *C. botulinum* could be isolated (25, 27). At most, honey exposure could account for only a minority of infant botulism cases, as just 35% of cases worldwide had been fed it prior to onset of illness (27). Nevertheless, honey is to date the one identified and avoidable source of *C. botulinum* spores to which susceptible infants have been exposed (24, 25, 27). The California Health Department, the Center for Disease Control, the Sioux Honey Association (the world's largest honey producer), and pediatric editorial opinion have all recommended that honey not be fed to any child 12 months of age or less (28).

Because milk is the principal food of infants, the possible role of breast milk and formula milk in predisposing to or protecting from illness was of particular interest. The interim data analysis was unable to provide a conclusion about the possible role of milk source because the number of patients in each of the feeding categories was too small to permit statistical testing (27). However, when more patients had become available for study, an association between milk source and severity of illness became evident (29). Dietary data were collected on all 60 California cases of infant botulism [50 hospitalized and 10 indistinguishable at autopsy from typical sudden infant death syndrome (SIDS)]. Of the hospitalized cases, 34 (68%) were primarily breast fed and 16 (32%) were primarily formula fed during the month prior to onset of illness. Of the healthy control infants for the corresponding time period, 42.4% were primarily breast fed and 57.6% were primarily formula fed. Unexpectedly, all ten cases of sudden infant death were primarily fed formula with added iron; no SIDS-like case of infant botulism occurred in any breast-fed infant. Hospitalized formula-fed infants were significantly younger at onset of illness than were breast fed infants. Breast-fed infants given supplementary iron became ill at a younger age than those breast-fed infants who did not receive iron. From these data it was concluded that breast feeding affords relative protection against infant botulism and allows the illness to evolve so gradually that hospitalization is possible, thus explaining the disproportionate number of hospitalized infants who were breast fed. In contrast, formula-feeding appears to predispose to fulminant, fatal SIDS-like illness. Supplementary iron in the diet, either from formula or from other sources may additionally contribute to susceptibility to the fulminant form of infant botulism (29).

The interim analysis of the California data found no association between illness and exposure to solid foods (27). However, since infants generally are introduced to solid foods in a defined progression, further analysis of risk by type of solid food exposure (e.g. cereals, vegetables, fruits) is underway.

Honey was not the only potential environmental source of *C. botulinum* spores to be identified. Spores were also found in a sample of vacuum cleaner dust collected at one patient's home, and in five soil samples, three obtained from the yards of case homes and two from the yards of control homes. As with honey, the toxin type (A or B) of the *C. botulinum* isolated from these environmental reservoirs matched in every case the toxin type of the patient's illness (27, 30). It deserves emphasis that for the majority of cases, no source of *C. botulinum* spores was identified, even circumstantially.

CLINICAL SPECTRUM

Most microbial pathogens are known to produce illness that in a population has a spectrum of severity usually ranging from the subclinical to the fulminant. It was therefore anticipated that infant botulism would similarly have a spectrum of clinical severity (31), even though all the early cases were sufficiently ill to need hospitalization (1, 7–9). Special study of groups other than those hospitalized provided support for this concept. A few infants had inapparent infection (transient "carriers" of *C. botulinum*), others had only mild, outpatient illness (31), and finally, at the most severe end of the spectrum, there were those infants who presented as cases of sudden, unexpected death that by history and autopsy were indistinguishable from typical cases of SIDS (crib death) (11, 31, 32). Because systematic studies of the mild and fulminant ends of the spectrum have to date been limited, most cases have been recognized in those infants whose onset of illness was gradual, yet whose course was sufficiently severe to require hospital admission.

Mild and Fulminant Illness

OUTPATIENTS Before infant botulism was known to exist as a distinct disease entity, mild cases were managed successfully under a variety of diagnostic labels and received a diagnostic evaluation that did not include botulism testing (19). In the spring of 1978 pediatricians in Utah who had had experience with hospitalized cases of infant botulism subsequently recognized mild cases that were managed as outpatients. Both patients had type A disease, and both had been considered cases of "failure to thrive" because for about three weeks they had failed to nurse adequately and gain weight. Supplementary formula feedings were carefully given at home and over the next several weeks the infants recovered their strength. During recovery, the diagnosis was suspected and confirmed by fecal examination (31).

The transient presence of *C. botulinum* in the feces of apparently healthy infants has been observed both in California and Utah. In the course of a screening program in California to determine the prevalence of *C. botulinum* in the fecal flora of normal infants, *C. botulinum* type A organisms but no botulinal toxin were identified in one of 160 infants studied. This eight-day-old infant developed constipation (but no other neurologic signs) about three weeks after the positive fecal specimen was passed, and the constipation responded to treatment with suppositories. Three subsequent fecal specimens collected over a six-week period were negative for *C. botulinum* organisms and toxin (11).

In Utah, a healthy 89-day-old baby was studied because his feces suddenly changed in color and odor from "yellow and cheese-like" to "green, hard, and foul-smelling." At no time was the infant constipated, and a day later the feces returned to their original character. The green, foul-smelling fecal specimen contained *C. botulinum* type A organisms and type A botulinal toxin. A second fecal specimen was tested three months later and did not contain either *C. botulinum* organisms or botulinal toxin (31). These observations illustrate that the presence of *C. botulinum* toxin and organisms in the bowel lumen does not always result in clinically-evident disease and provide a human parallel to similar experimental findings in infant mice (14).

The extent of mild and asymptomatic infection with *C. botulinum* will not be fully known until serologic tests to demonstrate its occurrence can be developed. More sensitive tests will also be needed to determine how much infant botulism occurs as fulminant disease that is indistinguishable from SIDS.

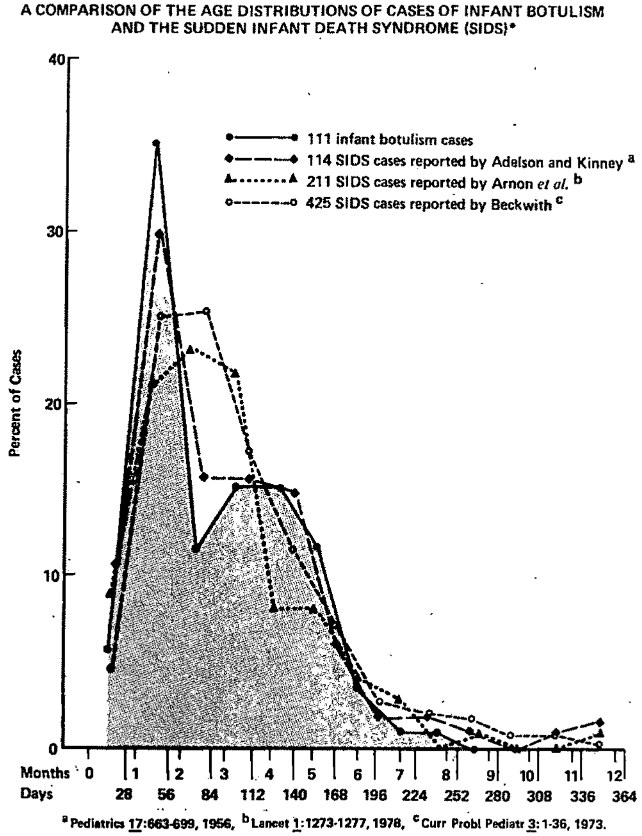
SUDDEN INFANT DEATH SYNDROME The relation of infant botulism to SIDS is of special interest. Present information on this relation derives from a 1977 California study of infant death that was undertaken once it became evident that botulinal toxin could be produced *in vivo* in the infant gut. The known potency and mechanism of action of the toxin suggested that its production could lead to rapid paralysis of airway and respiratory muscles that might result in sudden unexpected death (1, 9, 11, 31).

During 1977, autopsy specimens from 280 California infants who died from a variety of causes were tested for botulism. From 10 infants, all of whom died suddenly and unexpectedly, *C. botulinum* organisms were isolated. In two of the ten cases, botulinal toxin was identified as well. Nine of the deaths had been ascribed to SIDS by the forensic pathologist; the tenth had been called "cardiopulmonary failure." The histories and autopsy findings of the ten cases that contained *C. botulinum* were entirely typical of classical SIDS; a few infants had an antecedent upper respiratory infection, and at autopsy some had petechiae on intrathoracic organs, frothy intraalveolar fluid, or simply no lesions at all. The nine SIDS cases from which *C. botulinum* was isolated represented 4.3% of the SIDS cases studied, but because of various technical considerations, this figure appears to represent a minimum estimate of SIDS deaths caused by the intestinal production of botulinal toxin. In the control group of 68 infants who died of identified causes, neither *C. botulinum* organisms nor botulinal toxin was found (11). Subsequently, *C. botulinum* has also been isolated from a SIDS case in Utah (31) and in Washington (32), but not from other control infants who died of known causes (32).

Another observation which suggests that infant botulism may play a substantial etiologic role in SIDS is the similarity of the age distributions of the two entities. The age distribution of SIDS has long been considered one of its most characteristic features (47). Peak incidence of SIDS occurs at two to four months of age, and only some 10% of cases are older than six months. The remarkable similarity of the age distribution of the first 62 recognized infant botulism cases to that of SIDS (11) was maintained as the number of cases identified worldwide increased to 111 (Figure 1) (31).

Intestinally produced botulinal toxin may cause sudden death in infants by weakening or paralyzing the muscles of the tongue and pharynx that are innervated by cranial nerves IX–XII and which are among the muscle groups first affected in botulism. Flaccidity of the tongue and pharyngeal muscles would impede air flow with resultant hypoxemia that may further weaken these muscles and lead to frank obstruction of the upper airway. The infant may be especially vulnerable to such obstruction during sleep when conscious muscle tone is relaxed. Tonkin has cogently described the anatomic features that make the infant airway particularly vulnerable to such occlusion (33).

Fulminant cases of infant botulism that are indistinguishable from SIDS cases may absorb botulinal toxin from the gut in larger quantities or more rapidly than occurs in the hospitalized and outpatient cases. Such differential absorption may be due to individual variation among infants or may depend on the site in the gut where spores germinate, multiply, and produce toxin, as toxin is absorbed much more efficiently from the small bowel than from the large (34). The absorption of botulinal toxin that results in charac-


A COMPARISON OF THE AGE DISTRIBUTIONS OF CASES OF INFANT BOTULISM AND THE SUDDEN INFANT DEATH SYNDROME (SIDS)*

[a] Pediatrics 17:663-699, 1956, [b] Lancet 1:1273-1277, 1978, [c] Curr Probl Pediatr 3:1-36, 1973.

Figure 1 Hospitalized patient aged 44 days with infant botulism. Note ptosis, disconjugate gaze, flaccid and expressionless face, loss of head control requiring a supporting hand (nurse out of picture), as well as the presence of a cardiac monitor lead.

teristic SIDS fatalities may occur so rapidly that the signs and symptoms of illness seen in the typical hospitalized case (constipation, listlessness, poor feeding, weak cry, etc) do not have time to become apparent.

The Hospitalized Patient

ONSET AND PROGRESSION OF SYMPTOMS Because infants are unable to describe their symptoms, the onset of illness can only be detected by careful observation. Almost invariably, the first indication of illness is constipation (defined as three or more days without defecation), although this sign is frequently overlooked by both parents and physicians. Usually, parents first notice that the child has become lethargic, listless, and sleepier than usual. Appetite is diminished, as are the suck and gag reflexes. A nursing mother may notice breast engorgement. The child's cry becomes weak and feeble, and spontaneous activity is diminished. Dysphagia often is evident as secretions drooling from the mouth.

When the descending paralysis has affected bulbar musculature and then has become generalized, loss of head control is prominent. Not infrequently, flaccidity of airway and respiratory muscles may cause respiratory arrest at home or in the hospital.

In mild cases, or in the early stages of illness, physical signs of infant botulism may be subtle and easily missed. Cranial nerve palsies and the fatigability of muscular function must be elicited by careful examination. Ptosis often is not evident unless the child's head is held erect. Decreased anal sphincter tone may provide a clue that the patient has a generalized neuromuscular disease. Typically, extensive flaccid paralysis becomes evident in two to three days (19, 31), but in some cases, marked weakness may develop within hours. Some physical signs that the author has found helpful in detecting early or mild illness may be seen in Figure 2 and are listed in Table 1.

When the muscle weakness and hypotonia have become generalized (and most noticeable as loss of head control), the baby appears overtly "floppy" and the need for hospital admission is unquestioned—to provide supportive care and to continue the search for the etiology of the illness. The similarity of this presentation to that of other illnesses common in infancy often results in failure to consider infant botulism among the admission diagnoses. Instead, septicemia, viral infection, dehydration, idiopathic hypotonia, and polyneuritis have been among the most common initial diagnoses (Table 2) (19).

COURSE Once admitted, the course of hospitalized patients varied. In some cases, the flaccidity and weakness continued to progress after admission, while in others it did not. In general, if the infant's weakness continued

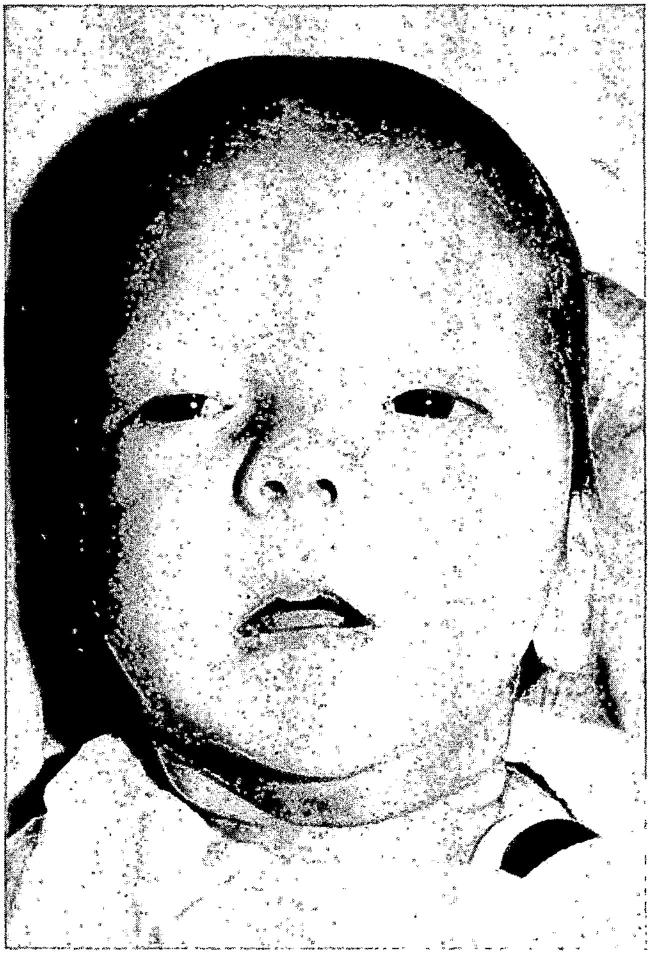


Figure 2 A comparison of the age distributions of cases of infant botulism and the sudden infant death syndrome. The solid line indicates all known infant botulism cases in the world: out-patient, in-patient, and sudden deaths. The age used was the age in days at death or when the infant was first seen by a physician for symptoms of botulism. ($N = 111$; 60 California, 49 other states, 1 England, 1 Australia.) [Reproduced with permission from Arnon, S. S., Chin, J. 1979. The clinical spectrum of infant botulism. *Rev. Infect. Dis.* 1:614–621. © 1979 by the University of Chicago.]

to extend after admission, it reached its maximum (which in most cases was well short of respiratory insufficiency) within the first week or two. Once a patient had reached his particular nadir, he usually remained at that level of weakness for two to three weeks before recovery of strength and movement became evident. Improvement was steady, but very slow, unless complications intervened.

At present, no reliable predictors of the clinical course of hospitalized patients have been identified. Illness resulting from type A toxin is potentially more severe than that resulting from type B toxin. Duration of hospitalization of California patients was significantly longer in cases caused by type A toxin than in those caused by type B toxin [mean of 39 vs 24 days, respectively (31)]. However, the range of hospital stay for cases caused by type A toxin was so large (4–112 days) that it encompassed the range for those caused by type B toxin (5–36 days); hence, the illness caused by type A toxin is not necessarily more severe than that caused by type B toxin.

Patients continued to excrete both *C. botulinum* toxin and organisms for weeks to months following recovery of sufficient strength to permit hospital discharge (35). In general, excretion of type A toxin and organisms lasted longer than did that of type B. Excretion of toxin and organisms for both type A and type B followed apparent first-order decay kinetics, and this fact

Table 1 Physical findings that may be helpful in the diagnosis of infant botulism

Test	Findings	Interpretation
1. Shine a bright light in the eye and observe the briskness of the pupillary light reflex. Once the iris constricts, remove the light. As soon as the iris dilates, again shine the light in the eye. Continue for 1–3 min without interruption.	Pupillary constriction, which may initially be brisk, eventually becomes sluggish. An initially mid-position pupil may become dilated at rest as the test continues.	Fatigability when repetitive muscle contraction is required is one hallmark of botulism. Infection—a most important consideration in the differential diagnosis at presentation—does not impair the pupillary light reflex (unless the infection is of the CNS).
2. Shine the light along the optical axis so that it falls directly on the fovea. Keep the light continuously on the fovea for 1–3 min even if the infant tries to deviate his eyes. Observe the infant's efforts and ability to deviate the eyes. Observe the vigor and purposefulness of the infant's efforts to avoid the light.	Latent ophthalmoplegia may be elicited. Efforts to push away the examiner's hand or to squirm away from the light may be feeble or not purposefully continued.	A sustained bright light on the fovea is most uncomfortable, and an infant with normal motor ability will use it to avoid this sensation.
3. Place a clean fifth finger in the infant's mouth (but be certain that doing so will not compromise adequate air flow). Note the strength and especially the duration of the reflex sucking efforts. If the infant has an empty stomach, advance the finger and manually gauge the briskness and strength of the gag reflex. (Feeling the gag reflex regularly is also a useful way to follow the infant's recovery.)	The suck is weak and of short duration. The gag reflex is diminished.	Fatigability, as in 1. above.

permitted calculation of the half-life of excretion. Substantial differences among the duration of excretion of type A and type B toxin and organisms were noted, but the reasons for these differences remain obscure. For type B toxin, median duration of excretion was 2.1 weeks, for type B organisms 3.0 weeks, for type A toxin 4.8 weeks, and for type A organisms 6.7 weeks. Excretion of *C. botulinum* organisms persisted 18 weeks after hospital discharge in one case. Thus, infant botulism is like other acute enteric bacterial infections in that, long after the episode of acute illness has resolved, fecal excretion of the pathogen persists (35). All patients eventually excluded *C. botulinum* from their fecal flora, an observation which emphasizes that *C. botulinum* is not part of the normal resident gut microflora of man (1, 9, 11, 36). The patients' recovery of strength despite the continued presence of botulinal toxin in their feces suggests either that the absorption of toxin ceased during recovery or that any toxin absorbed was inactivated before it reached the nerve endings.

COMPLICATIONS In retrospect, "progression" of "weakness" in the hospital (or "regression" once improvement began) has in some instances clearly been due to complications that have included nosocomial infection, intubation of a mainstem bronchus, hyponatremia (either from excess free water intravenously or from the syndrome of inappropriate antidiuretic hormone release), and tension pneumothoraces from a bronchopleural fistula. Also, necrotizing enterocolitis was reported in a patient with infant botulism (37).

Table 2 Diagnoses considered in patients whose illnesses were subsequently confirmed as infant botulism[a]

Initial impression at time of hospitalization	Common subsequent diagnoses
Sepsis	Disorders of amino acid metabolism
Viral syndrome	Metabolic encephalopathy
Dehydration	Hypothyroidism
Pneumonia	Heavy metal poisoning (lead, magnesium, arsenic)
Hypotonia of unknown etiology	Myasthenia gravis
Failure to thrive	Werdnig-Hoffman disease
	Drug/toxin ingestion
	Poliomyelitis
	Polyneuropathy secondary to viral infection
	Guillain-Barré syndrome or idiopathic infantile polyneuritis
	Hirschsprung's disease

[a] Modified from Johnson, R. O., Clay, S. A., Arnon, S. S. 1979. Diagnosis and management of infant botulism. *Am. J. Dis. Child.* 133:586-93.

MANAGEMENT The present treatment of infant botulism consists of meticulous supportive care, with particular attention directed to nutrition and to pulmonary hygiene. Immediate access to an intensive care unit and to mechanical ventilation is especially important because aspiration or apnea may easily occur. Monitoring for both apnea and bradycardia should be instituted and should continue until breathing, coughing, and swallowing ability have adequately returned.

The level of hospital care required was directly related to the extent of the patient's paralysis. For instance, one very mildly affected patient managed to continue to nurse successfully throughout hospitalization with the mother manually assisting and expressing breast milk into the baby's mouth. Most often, patients needed gavage feeding for at least part of their hospital stay. Several paralyzed patients required mechanical respirators to maintain life. About one third (15/50) of hospitalized California cases had a respiratory arrest after admission; 12 of these 15 patients needed at least some mechanical ventilatory support, but only two required tracheostomy. Occasionally, respiratory arrest was the consequence of simple mechanical factors, such as obstruction of the airway by pooled secretions or by improper positioning of the head and neck that allowed the flaccid tongue or pharyngeal muscles to occlude the airway. Correction of these oversights restored respiration. The other patients experienced respiratory insufficiency because muscular paralysis prevented their keeping the airway patent or kept the diaphragm from contracting effectively.

Tube feeding may stimulate peristalsis and has been successfully used in most patients, often utilizing mother's expressed breast milk (which has been of psychologic and physiologic benefit to her as well). Patients should not be fed by mouth until they are able to gag and swallow. To reduce the quantity of *C. botulinum* organisms and toxin in the intestine, cathartic agents or bulk laxatives may be judiciously administered if adynamic ileus is absent, but rarely have these proved efficacious. Repeated purgation is unwarranted and potentially hazardous. Since it is not known whether the *C. botulinum* organisms that are present in feces in large numbers [about 10^7/g (38)] may be infectious for other infants or adults by fecal-oral transmission, it is recommended that the usual precautions against fecal bacterial pathogens be taken during hospitalization. As some patients excrete *C. botulinum* toxin and organisms in their feces for weeks to months after they have returned home, the importance of careful handwashing and diaper disposal deserves emphasis. Botulinal toxin also causes bladder atony with resultant retained urine, and the risk of infection may be reduced by use of the Credé method. While recovering, infant botulism patients fatigue easily when sustained muscular action is required, an important consideration with resumption of oral feeding because of the possibility of aspiration.

That successful management of patients does not involve the use of botulinal antitoxin merits comment. Antitoxin is used in food-borne botulism only to neutralize circulating toxin, as botulinal toxin is irreversibly bound to nerve endings. Circulating toxin has not yet been found in patients with infant botulism with one exception. The exception was an Arizona case that occurred in September 1978 (39); the patient was not treated with antitoxin and recovered completely.

In California, antitoxin was not given to the early cases because the patients were in stable condition or had recovered by the time the diagnosis was established (1, 2, 7, 8). The experience gained with these and other patients indicated that antitoxin was not needed for complete recovery. Because botulinal antitoxin is a horse serum product, it may induce lifelong hypersensitivity (particularly if given to infants), and its use may also often be accompanied by substantial adverse side effects. When older children and adults with food-borne botulism were given antitoxin, serum sickness developed in 7.4% and anaphylaxis in 2.9% of recipients (40).

With two exceptions, botulinal antitoxin has not been given to any of the 126 patients hospitalized with infant botulism. The first exception was the ten-week-old New Jersey patient who in 1975 was thought to have food-borne botulism (1). The other exception was a 14-week-old Maryland patient who in 1978 was given antitoxin because paralysis and respiratory insufficiency progressed after hospital admission. Despite pretreatment with epinephrine, the second patient had an immediate anaphylactic reaction characterized by hypotension and bronchospasm after he had received intravenously only one-tenth the planned dose, the administration of which was promptly terminated. This case dramatizes the potential hazard of using equine botulinal antitoxin even in infants with no prior exposure to equine proteins (19, 31). Human-derived botulinal antitoxin would not have the limitations of the equine product, and the development of a human botulism immune globulin has begun (41).

As with antitoxin, the role of antibiotics—if any—in the treatment of infant botulism has yet to be determined. At present, several considerations argue against their use. Many patients received parenteral antibiotics, usually a penicillin derivative and an aminoglycoside, because sepsis was a commonly suspected diagnosis at admission. A few received oral penicillin or its derivatives, usually for secondary infections such as otitis media. When penicillin or its derivatives were used, and the feces examined at intervals subsequently, neither oral nor parenteral administration succeeded in producing discernible clinical benefit or in eradicating either *C. botulinum* organisms or botulinal toxin from the intestine (1, 19, 35). Also, it appears that in some instances parenteral aminoglycoside antibiotics may act in synergism with botulinal toxin at the neuromuscular junction and exacerbate the paralysis (42).

In addition, the use of oral antibiotics may have unforeseen consequences. Unlike most bacterial infections in which one species predominates, *C. botulinum* is just one of several clostridial and other anaerobic species present in the intestinal lumen of patients (38). Oral antibiotics that are clostridiocidal or clostridiostatic may be expected to alter the intestinal microecology in an unpredictable manner and might actually permit intestinal overgrowth by *C. botulinum.* Also, oral antibiotics that are efficiently absorbed higher in the gut might not reach the parts of the intestine where *C. botulinum* resides. It is even conceivable that the use of clostridiocidal antibiotics might substantially increase the amount of toxin available for absorption in the intestinal lumen, as botulinal toxin is liberated mainly when vegetative cell death and lysis occur (43). Despite these potential drawbacks, the possible merit of human antitoxin and of oral antibiotics in the therapy of infant botulism deserves to be explored in experimental animals before definite conclusions are drawn about their role in treatment.

Guanidine, a drug sometimes used in food-borne botulism, has not been given to patients with infant botulism. Even in food-borne botulism the efficacy of guanidine has not been convincingly demonstrated (44).

DIAGNOSIS AND DIFFERENTIAL DIAGNOSIS Although infant botulism is a long extant disease (2), physicians have heretofore mistaken it for other illnesses (19). (See Table 2.) These categories and an additional 17 less frequently considered possibilities suggest where presently occurring but unrecognized cases of infant botulism may be found (19).

If a previously healthy infant, usually six months of age or less, has a history of constipation and then develops weakness, with difficulty in sucking, swallowing, crying, or breathing, infant botulism should be considered the most likely diagnosis (1, 7–9, 19). Electromyography (EMG) may be quite helpful at the bedside, as a characteristic pattern known as Brief, Small, Abundant motor-unit action Potentials (BSAP) has frequently been found, but its absence does not exclude the diagnosis of infant botulism (1, 7, 8, 19). The EMG pattern of post-tetanic facilitation has also been observed in patients with infant botulism (7, 8, 45).

The diagnosis of infant botulism is established only by the identification of *C. botulinum* organisms in the feces or in other specimens collected at autopsy (1, 2, 9, 11, 32). If a fecal specimen of adequate size can be obtained (the constipation can sometimes make this task difficult), botulinal toxin can invariably be found in it as well. Study of the feces of 160 healthy infants age-matched to cases of infant botulism demonstrated that *C. botulinum* organisms are not part of the normal resident intestinal microflora of infants (11). As previously mentioned this conclusion was buttressed by the observation that when the fecal excretion of *C. botulinum* organisms and toxin of 38 patients with infant botulism was followed, all patients eventually eliminated both the toxin and organisms from their feces (35). Examination of the serum for the presence of botulinal toxin cannot be relied on to make the diagnosis of infant botulism because, with the one exception noted (39), no circulating toxin has been found in the sera of any patient.[1]

PROGNOSIS For those infant botulism cases whose onset of illness is sufficiently gradual to permit hospitalization, prognosis is excellent and recovery complete. Gradual recovery of strength occurs through sprouting of new terminal motoneurons that locate noncontracting muscle fibers and then reinnervate them by inducing formation of a new motor end-plate (46). Clinically and in experimental animals, this process takes some weeks to be completed (1, 19, 46).

Only 3 of the 126 patients with laboratory-confirmed infant botulism that have been admitted to hospitals worldwide have died, one from nosocomial complications and the other two from cerebral hypoxia that followed unexpected respiratory arrest at home. The case-fatality ratio of hospitalized cases of infant botulism may be expected to be substantially higher than 2.4% in countries that have limited intensive care facilities.

It has been suggested that recovery occurs in infant botulism in part through development of protective antibody (1), but it has not been possible to test for an antibody response with the currently available mouse neutralization test because the test requires larger volumes of serum than can be safely collected from ill infants. With the development of more sensitive assays, e.g. an enzyme-linked immunosorbent assay (ELISA) system, it may be possible to determine whether there is a humoral or secretory antibody response, or both, to *C. botulinum* infection.

SUMMARY: UNANSWERED QUESTIONS

The recognition of infant botulism as an infectious disease and its subsequent investigation have served to clarify the areas in which more information is needed. Important epidemiologic, clinical, and pathophysiologic questions await answers.

In order to determine the true incidence of infant botulism, the proportion of cases that occur in the mild (outpatient), moderate to severe (hospitalized), and fulminant (SIDS) portions of the clinical spectrum needs to be ascertained. This, however, requires widespread awareness of the disease and access to diagnostic laboratories. Once these conditions have been fulfilled, it may be possible to learn whether there is geographic variation in incidence and in seasonality, and if so, to probe for an explanation. Other vehicles of *C. botulinum* spores besides honey and the routes by which such spores reach the infant gut await identification.

---

[1] Several State Health Department laboratories are equipped to test infant botulism specimens. Physicians wishing this diagnostic examination should contact their State Health Department for specimen testing or for possible referral to the Center for Disease Control (CDC), Atlanta, Georgia. Physicians who suspect food-borne botulism (a medical and public health emergency) in older children or adults are urged to alert their State Health Department and the CDC (phone 404-329-3311 days, 404-329-3644 nights, holidays, and weekends).

At present, the most important clinical questions concern treatment. Meticulous supportive care is effective, but is improvement possible? For instance, can hospital stay be shortened or severity of illness ameliorated by use of selected antibiotics or of human-derived botulinal antitoxin?

The answers to some of the most clinically germane questions may be had by gaining further insight into the pathogenesis of infant botulism. The ubiquitous distribution of *C. botulinum* spores implies that exposure to them is probably a universal, but intermittent, occurrence. Why then do the spores germinate in the intestines of only some infants? And among those infants in whom the spores multiply, why is the course of illness mild in some children and fulminant in others? To what extent is clinical severity the consequence of host factors, such as rate of absorption of toxin, and to what extent is severity a consequence of agent factors, such as the rate of toxin production by different *C. botulinum* strains? The answers to these questions will probably be found in a quantitative understanding of the roles of host physiology and immunity (including that derived from the mother), of diet in determining the composition of the intestinal microflora, and of the possible synergistic and antagonist effects of the gut microflora on *C. botulinum* spore germination and subsequent toxin production.

There remains the riddle of the restricted age of occurrence of infant botulism and its peculiar distribution, which is so strikingly similar to that of SIDS. If the answers to the pathophysiology of infant botulism explain its age distribution, then they also may provide insight into a pathogenesis common to the presumed many etiologies of SIDS. Finally, infant botulism, including its SIDS component, may eventually be prevented by applying measures used to control other infectious diseases. How soon this can be accomplished is the remaining question.

Literature Cited

1. Arnon, S. S., Midura, T. F., Clay, S. A., Wood, R. M., Chin, J. 1977. Infant botulism: epidemiological, clinical, and laboratory aspects. *J. Am. Med. Assoc.* 237:1946–51
2. Arnon, S. S., Werner, S. B., Faber, H. K., Farr, W. H. 1979. Infant botulism in 1931: discovery of a misclassified case. *Am. J. Dis. Child.* 133:580–82
3. Young, J. H. 1976. Botulism and the ripe olive scare of 1919–1920. *Bull. Hist. Med.* 50:372–91
4. Geiger, J. C., Dickson, E. C., Meyer, K. F. 1922. The epidemiology of botulism. *Public Health Bull.* 127:1–119
5. Grover, W. D., Peckham, G. J., Berman, P. H. 1974. Recovery following cranial nerve dysfunction and muscle weakness in infancy. *Dev. Med. Child Neurol.* 16:163–71
6. Ramseyer, J. C., Clay, S. A., Fishman, L. S. 1976. Electromyographic studies in acute infantile polyneuropathy. *Neurology* 26:364
7. Clay, S. A., Ramseyer, J. C., Fishman, L. S., Sedgwick, R. P. 1977. Acute infantile motor unit disorder: infantile botulism? *Arch. Neurol.* 34:236–43
8. Pickett, J., Berg, B., Chaplin, E., Brunstetter-Shafer, M. 1976. Syndrome of botulism in infancy: clinical and electrophysiologic study. *N. Engl. J. Med.* 295:770–72
9. Midura, T. F., Arnon, S. S. 1976. Infant botulism: identification of *Clostridium botulinum* and its toxins in faeces. *Lancet* 2:934–36
10. Smith, L. DS. 1977. *Botulism: the Organism, its Toxins, the Disease,* Springfield, Ill: Thomas
11. Arnon, S. S., Midura, T. F., Damus, K., Wood, R. M., Chin, J. 1978. Intestinal infection and toxin production by *Clostridium botulinum* as one cause of Sudden Infant Death Syndrome. *Lancet* 1:1273–77
12. Orr, P. F. 1922. The pathogenicity of *Bacillus botulinus. J. Infect. Dis.* 30:118–27
13. Miyazaki, S., Sakaguchi, G. 1978. Experimental botulism in chickens: the cecum as the site of production and absorption of botulinum toxin. *Jpn. J. Med. Sci. Biol.* 31:1–15
14. Sugiyama, H., Mills, D. C. 1978. Intraintestinal toxin in infant mice challenged intragastrically with *Clostridium botulinum* spores. *Infect. Immun.* 21:59–63
15. Mills, D. C., Sugiyama, H. 1979. *Comparative sensitivities of infant and adult mice to botulinum toxin.* Ann. Meet. Am. Soc. Microbiol., 79th, Los Angeles. Abstract #P11
16. Moberg, L. J., Sugiyama, H. 1979. Microbial ecologic basis of infant botulism as studied with germfree mice. *Infect. Immunol.* 25:653–57
17. Turner, H. D., Brett, E. M., Gilbert, R. J., Ghosh, A. C., Liebeschuetz, H. J. 1978. Infant botulism in England. *Lancet* 1:1277–78
18. Shield, L. K., Wilkinson, R. G., Ritchie, M. 1978. Infant botulism in Australia (letter). *Med. J. Aust.* 2:157
19. Johnson, R. O., Clay, S. A., Arnon, S. S. 1979. Diagnosis and management of infant botulism. *Am. J. Dis. Child.* 133:586–93
20. Meyer, K. F., Dubovsky, B. J. 1922. The distribution of the spores of *B. botulinus* in the United States. *J. Infect. Dis.* 31:559–94
21. Smith, L. DS. 1978. The occurrence of *Clostridium botulinum* and *Clostridium tetani* in the soil of the United States. *Health Lab. Sci.* 15:74–80
22. Thompson, J. A., Santos, J. I., Glasgow, L. A. 1979. Infantile botulism: spectrum of clinical presentation and epidemiology. *Clin. Res.* 27:115A
23. Follow-up on infant botulism—United States. 1978. *Morbid. Mortal. Wkly. Rep.* 27:17–23
24. Midura, T. F., Snowden, S., Wood, R. M., Arnon, S. S. 1979. Isolation of *Clostridium botulinum* from honey. *J. Clin. Microbiol.* 9:282–83
25. Midura, T. F. 1979. Laboratory aspects of infant botulism in California. *Rev. Infect. Dis.* 1:652–54
26. Cohn, V. 1978. Risk of botulism seen in feeding honey to infants under one year. *Washington Post,* July 6, p. 1
27. Arnon, S. S., Midura, T. F., Damus, K., Thompson, B., Wood, R. M., Chin, J. 1979. Honey and other environmental risk factors for infant botulism. *J. Pediatr.* 94:331–36
28. Brown, L. W. 1979. Infant botulism and the honey connection (commentary). *J. Pediatr.* 94:337–38
29. Arnon, S. S., Damus, K., Thompson, B., Midura, T. F., Chin, J. 1979. *Formula feeding and dietary iron as risk factors for sudden infant death syndrome caused by botulinal toxin.* Abstract No. 1048. *11th Int. Congr. Chemother.—19th Intersci. Conf. on Antimicrobial Agents and Chemotherapy,* Boston, Mass: Am. Soc. Microbiol.

30. Chin, J., Arnon, S. S., Midura, T. F. 1979. Food and environmental aspects of infant botulism in California. *Rev. Infect. Dis.* 1:693–96
31. Arnon, S. S., Chin, J. 1979. The clinical spectrum of infant botulism. *Rev. Infect. Dis.* 1:614–21
32. Peterson, D. R., Eklund, M. W., Chinn, N. M. 1979. The Sudden Infant Death Syndrome and infant botulism. *Rev. Infect. Dis.* 1:630–34
33. Tonkin, S. 1975. Sudden Infant Death Syndrome: hypothesis of causation. *Pediatrics* 55:650–61
34. Dack, G. M., Hoskins, D. 1942. Absorption of botulinum toxin from the colon of *Macaca mulatta. J. Infect. Dis.* 71:260–63
35. Arnon, S. S., Midura, T. F., Damus, K., Snowden, S., Thompson, B., Chin, J. 1979. *Persistent fecal excretion of Clostridium botulinum toxin and organisms in patients recovering from infant botulism.* Ann. Meet. Am. Soc. Microbiol., 79th, Honolulu, Hawaii. Abstract No. C(H)67
36. Dowell, V. R. Jr., McCroskey, L. M., Hatheway, C. L., Lombard, G. L., Hughes, J. M., Merson, M. H. 1977. Coproexamination for botulinal toxin and *Clostridium botulinum. J. Am. Med. Assoc.* 238:1829–32
37. Edmond, B. J., Guerra, F. A., Blake, J., Hempler, S. 1977. Case of infant botulism in Texas. *Tex. Med.* 73:85–88
38. Wilcke, B. W. Jr., Midura, T. F., Arnon, S. S. Quantitative evidence of gut colonization by *Clostridium botulinum* in four cases of infant botulism. *J. Infect. Dis.* In press
39. Alexander, D., Kaplan, A., Lersch, A., Kelter, A. 1978. Infant botulism—Arizona. *Morbid. Mortal. Wkly. Rep.* 27:411–12
40. Merson, M. H., Hughes, J. M., Dowell, V. R., Taylor, A., Barker, W. H., Gangarosa, E. J. 1974. Current trends in botulism in the United States. *J. Am. Med. Assoc.* 229:1305–8
41. Metzger, J. F., Lewis, G. E. Jr. 1979. Human-derived immune glubulins for the treatment of botulism. *Rev. Infect. Dis.* 1:689–90
42. L'Hommedieu, C., Stough, R., Brown, L., Kettrick, R., Polin, R. 1979. Potentiation of neuromuscular weakness in infant botulism by aminoglycosides. *J. Pediatr.* 95:1065–70
43. Bonventre, P. F., Kempe, L. L. 1960. Physiology of toxin production by *Clostridium botulinum* types A and B. I. Growth, autolysis, and toxin production. *J. Bacteriol.* 79:18–23
44. Faich, G. A., Graebner, R. W., Sato, S. 1971. Failure of guanidine therapy in botulism A. *N. Engl. J. Med.* 285:773–76
45. Snyder, R. D., Bicknell, J. M. Enhancement of evoked electromyographic response in infant botulism (letter). *Am. J. Dis. Child.* In press
46. Duchen, L. W. 1972. Motor nerve growth induced by botulinum toxin as a regenerative phenomenon. *Proc. R. Soc. Med.* 65:196–97
47. Peterson, D. R., van Belle, G., Chinn, N. M. 1979. Epidemiological comparisons of the Sudden Infant Death Syndrome with other major components of infant mortality. *Am. J. Epidemiol.* 110:699–707

CUMULATIVE INDEXES

CONTRIBUTING AUTHORS, VOLUMES 27–31

A

Abildskov, J. A., 30:171–79
Ahmed, S. S., 28:393–409
Ahn, Y. S., 28:299–309
Alexander, P., 27:207–24
Alfrey, A. C., 29:93–98
Annas, G. J., 29:9–14
Anthony, B. F., 28:355–69
Arnold, A. B., 28:187–94
Arnon, S. S., 31:541–60
Athanasoulis, C. A., 30:41–59
Atkinson, J. P., 28:291–98

B

Bailey, J. C., 29:417–26
Baldwin, D. S., 27:49–55
Baldwin, R. W., 27:151–63
Barnes, A. C., 27:1–7
Barnes, R. W., 31:201–18
Begemann, H., 28:131–41
Bergsagel, D. E., 30:431–43
Betts, R. F., 28:103–10
Bierman, E. L., 30:155–70
Bigger, J. T. Jr., 30:91–118
Blankenhorn, D. H., 30:289–300
Bollet, A. J., 31:463–87
Bonadonna, G., 29:149–75
Bond, J. H., 31:127–37
Bone, H. G. III, 28:111–17
Bouchier, I. A. D., 31:59–77
Boyar, R. M., 29:509–20
Branch, R. A., 29:205–18
Braylan, R. C., 30:17–24
Brenner, B. M., 28:91–102
Bresnick, G. H., 30:331–38
Bressler, R., 31:329–43
Bristow, J. D., 29:475–83
Burke, J. S., 30:17–24
Buselmeier, T. J., 31:345–60

C

Cannon, P. J., 27:485–521
Cantor, H., 30:269–77
Caro, W. A., 28:251–69
Cavalieri, R. R., 28:57–65
Cerami, A., 31:29–34
Cobb, L. A., 31:453–62
Cohen, J. J., 29:521–41
Coleman, T. G., 31:15–27
Colman, N., 31:433–39
Colman, R. W., 30:359–74
Colvin, R. B., 29:191–203
Conklin, R. H., 28:167–77
Cooper, T. B., 30:259–68
Copass, M. K., 31:453–62
Craig, W. A., 27:287–300
Crawford, D. W., 30:289–300
Crosnier, J., 29:67–72

D

D'Angio, G. J., 30:405–15
Dean, R. H., 27:397–405
DeClue, J. W., 31:15–27
Dement, W. C., 27:465–84
Douglas, R. G. Jr., 31:233–38
Droegemueller, W., 31:329–43
Dudrick, S. J., 28:517–28
DuPont, H. L., 28:167–77
Durrer, D., 27:63–68
Dustan, H. P., 29:485–93

E

Ehrhardt, A. A., 30:417–30
Eickhoff, T. C., 28:411–23
Elharrar, V., 29:417–26
Elin, R. J., 27:127–41
Ellis, A. K., 31:489–508
Ensinck, J. W., 30:155–70
Epstein, C. J., 29:117–28
Ettinger, P. O., 28:393–409
Ettlinger, R. E., 29:15–22
Evenson, M. A., 28:85–90
Everett, R. B., 30:473–88

F

Fajans, S. S., 30:313–29
Fanelli, G. M. Jr., 28:349–54
Fang, L. S. T., 30:225–39
Feinstone, S. M., 29:359–66
Fialkow, P. J., 30:135–43
Finegold, S. M., 31:219–32
Fishman, A. P., 31:421–31
Fitzgerald, F., 29:177–89
Floyd, J. C. Jr., 30:313–29
Foreman, J. C., 31:181–90
Forstot, S. L., 28:21–35
Frank, M. M., 28:291–98
Freeman, H. J., 29:99–116
Freeman, R. G., 27:77–87
Freis, E. D., 30:81–89
Friedman, E. A., 31:463–87
Friedman, R. C., 27:57–62
Froelicher, V. F. Jr., 28:1–12

G

Gasparini, M., 29:149–75
Gault, M. H., 27:117–25
Gavras, H., 27:485–521
Gennari, F. J., 29:521–41
Gerald, P. S., 29:563–70
Gerich, J. E., 27:379–88
Giebink, G. S., 29:285–306
Gill, J. R. Jr., 31:405–19
Gillis, R. A., 29:73–79
Ginzler, E. M., 31:463–87
Glatstein, E., 29:81–91
Gluck, F. W., 27:397–405
Golbus, M. S., 29:117–28
Goldberg, H. I., 30:181–88
Goldfine, I. D., 29:407–15
Goldsmith, R. S., 27:181–90
Goodnight, J. E. Jr., 29:231–83
Graber, S. E., 29:51–66
Grams, R. R., 27:199–206
Green, R., 27:57–62
Greenblatt, D. J., 27:407–20
Greer, M. A., 31:385–404
Grossman, M., 28:43–55
Grossman, M. I., 29:495–507
Grumet, F. C., 28:425–52
Guillemin, R., 27:379–88
Guilleminault, C., 27:465–84
Gutterman, J. U., 28:489–515
Guyton, A. C., 31:15–27

H

Habener, J. F., 29:327–42
Hackett, T. P., 27:371–78

Haider, B., 28:393–409
Hall, C. L., 30:199–211
Hall, W. J., 31:233–38
Hallstrom, A. P., 31:453–62
Halmi, K. A., 29:137–48
Halsted, C. H., 31:79–87
Hamburger, J., 29:67–72
Hanshaw, J. B., 28:103–10
Hardwicke, J., 30:199–211
Harker, L. A., 31:361–71
Harrington, W. J., 28:299–309
Hartmann, R. C., 28:187–94
Havel, R. J., 28:195–209
Herbert, V., 31:433–39
Herndon, J. H. Jr., 27:77–87
Hersh, E. M., 28:489–515
Hetzel, D., 30:61–79
Holland, J. F., 28:475–88
Hollister, L. E., 27:431–46
Holmsen, H., 30:119–34
Hook, E. W., 28:371–91
Hopewell, P. C., 27:343–56
Hornick, R. B., 30:457–72
Hoyumpa, A. M. Jr., 29:205–18
Hunder, G. G., 29:15–22
Hurst, M. W., 27:301–12
Hutchins, G. M., 31:289–301

I

Irsigler, G. B., 31:109–26

J

Janowitz, H. D., 27:269–85
Jenkins, C. D., 27:301–12; 29:543–62
Jordan, P. H. Jr., 27:191–97

K

Kadir, S., 30:41–59
Kales, A., 30:301–12
Kales, J. D., 30:301–12
Kaplan, B. S., 27:117–25
Kaplan, N. M., 29:31–40
Karnovsky, M. J., 30:213–24
Kattus, A. A., 27:69–75
Kaufman, H. E., 28:21–35
Kawai, C., 31:139–57
Kawamura, K., 31:139–57
Kaye, M. D., 30:375–404
Kim, Y., 29:319–25; 31:273–88
Kim, Y. S., 29:99–116
Kinlough-Rathbone, R. L., 31:89–96
Kinney, D. K., 29:459–73
Kitabchi, A. E., 30:339–57
Kitzmiller, J. L., 30:375–404
Kjellstrand, C. M., 31:345–60
Klassen, J., 27:117–25
Klein, M. S., 27:89–108
Klein, P. A., 28:311–27
Klocke, F. J., 31:489–503
Knochel, J. P., 30:145–53
Koenig, R. J., 31:29–34
Korenman, S. G., 27:357–70
Korobkin, M., 30:181–88
Krantz, S. B., 29:51–66
Kunin, C. M., 27:287–300

L

Lamberti, J. J., 29:571–82
Langer, G. A., 28:13–20
Levitt, M. D., 31:127–37
Lewis, E. J., 30:445–55
Lichtenstein, L. M., 31:181–90
Linnemann, C. C. Jr., 28:179–85
Lipscomb, K., 28:271–89
Littman, P. S., 30:405–16
Lockwood, C. M., 31:167–79
Lofgren, S., 27:407–20
Logue, G. L., 31:191–200
Lohmeier, T. E., 31:15–27
Long, J. M. III, 28:517–28
Lopez, M., 27:453–63
Love, R., 28:131–41
Lyons, M. M., 28:393–409

M

Maas, J. W., 29:219–29
MacDonald, P. C., 30:473–88
MacLean, W. A. H., 30:259–68
Maddox, D. A., 28:91–102
Mahaffey, J. E., 29:327–42
Majerus, P. W., 29:41–49
Malarkey, W. B., 30:245–58
Mandel, L. P., 31:453–62
Marchesi, V. T., 29:593–603
Marcus, R., 27:357–70
Matsumori, A., 31:139–57
Matthysse, S., 29:459–73
Mavligit, G. M., 28:489–515
McCabe, W. R., 27:335–41
McClung, M. R., 31:385–404
McCluskey, R. T., 29:191–203
McDevitt, H. O., 28:425–52
McGuigan, J. E., 29:307–18
McPhaul, J. J. Jr., 28:37–42
McRee, D. I., 27:109–15
Meares, E. M. Jr., 30:275–88
Mello, N. K., 27:321–33
Mendelson, J. H., 27:321–33
Merin, R. G., 28:75–83
Merrill, J. P., 29:343–58
Meyer, R. D., 31:219–32
Meyer-Bahlburg, H. F. L., 30:417–30
Michael, A. F., 29:319–25; 31:273–88
Miletich, J. P., 29:41–49
Miller, G. J., 31:97–108
Millman, B. S., 28:223–3[illegible]
Minna, J. D., 30:359–74
Mitchel, D. H., 31:51–57
Modell, J. H., 29:1–8
Morris, L., 30:339–57
Morton, D. L., 29:231–83
Moss, A. J., 31:1–14
Moylan, J. A., 28:85–90
Mullins, C. B., 28:271–89
Murray, J. F., 27:343–56
Mustard, J. F., 31:89–96
Myers, F. L., 30:331–38

N

Nasenbeny, J., 28:179–85
Noël, L.-H., 29:67–72

O

Oates, J. A., 29:23–29
Odell, W. D., 29:379–406
Okada, D. M., 28:355–69
Oldewurtel, H. A., 28:393–409
Oliver, J. A., 27:485–521
Ontjes, D. A., 27:263–68
Orci, L., 28:119–30
O'Reilly, R. A., 27:245–61
Orne, M. T., 30:489–503
Osborn, A. G., 20:189–98

P

Packham, M. A., 31:89–96
Pak, C. Y. C., 28:111–17
Parisi, A. F., 28:159–66
Perloff, M., 28:475–88
Peschle, C., 31:303–14
Peters, D. K., 31:167–79
Pierce, A. K., 28:211–21
Pietra, G. G., 31:421–31
Pitot, H. C., 30:25–39
Plotkin, S. A., 29:583–91
Plumb, V. J., 30:259–68
Portlock, C. S., 29:81–91
Portnoy, B. L., 28:167–77
Price, M. R., 27:151–63
Purcell, R. H., 29:359–66

Q

Quest, J. A., 29:73–79
Quie, P. G., 29:285–306

R

Raisz, L. G., 29:23–29
Rakowski, H., 27:165–80
Ranganathan, N., 27:165–80
Rapoport, B., 28:57–65
Regan, T. J., 28:393–409
Reiffel, J. A., 30:91–118
Renfrew, R., 31:345–60
Resnekov, L., 29:571–82

Robboy, S. J., 30:359–74
Robertson, G., 31:315–27
Robertson, J., 28:211–21
Robins, E., 28:67–73
Rocklin, R. E., 30:375–404
Rose, R. M., 27:301–12
Rosenberg, R. D., 29:367–78
Ross, R., 30:1–15
Rossi, A., 29:149–75
Rubin, R. H., 30:225–39
Russell, F. E., 31:247–59
Ryan, W. G., 28:143–52

S

Sachar, D. B., 27:269–85
Sachar, E. J., 27:389–96
Sacks, H., 30:339–57
Salvaggio, J., 27:453–63
Sanford, J. P., 27:421–29
Sasahara, A. A., 28:159–66
Sasazuki, T., 28:425–52
Schacht, R. G., 27:49–55
Schenker, S., 29:205–18
Schiff, G. M., 27:447–52; 31:441–51
Scott, H. W. Jr., 27:397–405
Seltzer, H. S., 31:261–72
Severinghaus, J. W., 31: 109–26
Seyberth, H. W., 29:23–29
Shader, R. I., 27:407–20
Shah, P. M., 28:235–50
Shah, S. A., 29:563–70
Sharma, G. V. R. K., 28:159–66
Shearman, D. J. C., 30:61–79
Sherlock, S., 27:143–49
Sherwin, I., 27:37–47
Shimm, D. S., 31:191–200
Shull, H. J., 27:397–405
Silen, W., 27:9–22
Silver, M. D., 27:165–80
Sipes, J. N., 28:371–91
Skillman, J. J., 27:9–22
Slichter, S. J., 31:509–40
Smith, J. W., 31:373–83
Smith, R. T., 28:311–27
Snyder, W. H. III, 28:111–17
Sobel, B. E., 27:89–108
Soldatos, C. R., 30:301–12
Soll, A. H., 29:495–507
Spitzer, R. L., 27:57–62
Stephen, C. R., 28:153–57
Stossel, T. P., 29:427–57
Stropes, L., 31:315–27
Stryckmans, P. A., 31:59–66
Sweeney, D. R., 29:219–29

T

Tancini, G., 29:149–75
Tannen, R. L., 31:35–49
Tarazi, R. C., 29:485–93
Theml, H., 28:131–41
Thompson, R. G., 31:453–62
Thompson, R. L., 28:371–91
Tilkian, A., 27:465–84
Tolkoff-Rubin, N. E., 30:225–39
Tomasi, T. B. Jr., 28:453–65
Tow, D. E., 28:159–66

U

Unger, R. H., 28:119–30

V

Voorhees, J. J., 28:467–73

W

Waldo, A. L., 30:259–68
Wall, R. T., 31:361–71
Walzer, S., 29:563–70
Ward, L. E., 29:15–22
Warren, J. W., 30:457–72
Waterson, A. P., 27:23–35
Wechsler, A. S., 28:341–48
Weiss, H. J., 30:119–34
Weitzman, E. D., 27:225–43
Wells, J. L. Jr., 30:259–68
Wells, S. A. Jr., 27:263–68
Wessler, S., 27:313–19
Wigle, E. D., 27:165–80
Wiktor, T., 29:583–91
Wilfert, C. M., 29:129–36
Wilson, D. R., 28:329–39
Winkelmann, R. K., 28:251–59
Wisneski, J. A., 29:475–83
Wolfe, M. S., 31:373–83
Wolff, S. M., 27:127–41
Wolfsen, A. R., 29:379–406

Y

Yalow, R. S., 30:241–48
Young, D. B., 31:15–27
Young, R., 30:339–57

Z

Zerbe, R., 31:315–27
Ziegler, M. G., 31:239–45
Zipes, D. P., 29:417–26

CHAPTER TITLES, VOLUMES 27–31

ANESTHESIOLOGY

Effects of Anesthetic Drugs on Myocardial Performance in Man	R. G. Merin	28:75–83
Malignant Hyperpyrexia	C. R. Stephen	28:153–57
Acupuncture: Context and Critique	B. S. Millman	28:223–34

CARDIOVASCULAR SYSTEM

Medical and Surgical Treatment of the Wolff-Parkinson-White Syndrome	D. Durrer	27:63–68
Coronary Artery Spasm: A New Appraisal	A. A. Kattus	27:69–75
Medical Management of Myocardial Infarction	M. S. Klein, B. E. Sobel	27:89–108
Mitral Valve Prolapse	E. D. Wigle, H. Rakowski, N. Ranganathan, M. D. Silver	27:165–80
The Detection of Asymptomatic Coronary Artery Disease	V. F. Froelicher Jr.	28:1–12
Ionic Basis of Myocardial Contractility	G. A. Langer	28:13–20
Hypertrophic Cardiomyopathy	P. M. Shah	28:235–50
Development of Coronary Collateral Circulation	A. S. Wechsler	28:341–48
Prophylaxis of Infective Endocarditis: A Reevaluation	J. N. Sipes, R. L. Thompson, E. W. Hook	28:371–91
The Role of Ethanol in Cardiac Disease	T. J. Regan, P. O. Ettinger, B. Haider, S. S. Ahmed, H. A. Oldewurtel, M. M. Lyons	28:393–409
Neural Actions of Digitalis	R. A. Gillis, J. A. Quest	29:73–79
Slow-Channel Depolarization: Mechanism and Control of Arrhythmias	J. C. Bailey, V. Elharrar, D. P. Zipes	29:417–26
Left Ventricular Stiffness	J. A. Visneski, J. D. Bristow	29:475–83
Cardiogenic Hypertension	H. P. Dustan, R. C. Tarazi	29:485–93
Behavioral Risk Factors in Coronary Artery Disease	J. D. Jenkins	29:543–62
Cardiac Assist Devices	J. J. Lamberti, L. Resnekov	29:571–82
The Arterial Wall and Atherosclerosis	R. Ross	30:1–15
Validity of Therapy for Mild Hypertension	E. D. Freis	30:81–89
Sick Sinus Syndrome	J. T. Bigger Jr., J. A. Reiffel	30:91–118
The Prolonged QT Interval	J. A. Abildskov	30:171–79
Studies of Atrial Flutter Following Open Heart Surgery	A. L. Waldo, J. L. Wells Jr., V. J. Plumb, T. B. Cooper, W. A. H. MacLean	30:259–68
Regression of Atherosclerosis	D. W. Crawford, D. H. Blankenhorn	30:289–300
Prediction and Prevention of Sudden Cardiac Death	A. J. Moss	31:1–14
Salt Balance and Long-Term Blood Pressure Control	A. C. Guyton, T. G. Coleman, D. B. Young, T. E. Lohmeier, J. W. DeClue	31:15–27
High Density Lipoproteins and Atherosclerosis	G. J. Miller	31:97–108
Myocardial Biopsy	C. Kawai, A. Matsumori, K. Kawamura	31:139–57
Noninvasive Evaluation of the Carotid Bruit	R. W. Barnes	31:201–18

Postural Hypotension	M. G. Ziegler	31:239–45
Pathological Changes in Aortocoronary Bypass Grafts	G. M. Hutchins	31:289–301
Pericarditis and Renal Failure	R. Renfrew, T. J. Buselmeier, C. M. Kjellstrand	31:345–60
Primary Pulmonary Hypertension	A. P. Fishman, G. G. Pietra	31:421–31
Community Cardiopulmonary Resuscitation	L. A. Cobb, A. P. Hallstrom, R. G. Thompson, L. P. Mandel, M. K. Copass	31:453–62
Control of Coronary Blood Flow	F. J. Klocke, A. K. Ellis	31:489–508

CLINICAL PHARMACOLOGY

Rational Psychopharmacology for Patients with Medical Diseases	D. J. Greenblatt, R. I. Shader, S. Lofgren	27:407–20
Effective Use of Analgesic Drugs	L. E. Hollister	27:431–46
The Disposition and Effects of Sedatives and Analgesics in Liver Disease	A. M. Hoyumpa Jr., R. A. Branch, S. Schenker	29:205–18
Clinical Pharmacology of Acute Allergic Disorders	J. C. Foreman, L. M. Lichtenstein	31:181–90

DERMATOLOGY

Human Disease Associated with Exposure to Light	J. H. Herndon Jr., R. G. Freeman	27:77–87
Current Problems in Mycosis Fungoides and Sezary Syndrome	R. K. Winkelmann, W. A. Caro	28:251–69
Pathophysiology of Psoriasis	J. J. Voorhees	28:467–73

ENDOCRINOLOGY AND METABOLISM

Circadian Rhythms and Episodic Hormone Secretion in Man	E. D. Weitzman	27:225–43
Estrogens and the Human Male	R. Marcus, S. G. Korenman	27:357–70
Somatostatin: Physiological and Clinical Significance	R. Guillemin, J. E. Gerich	27:379–88
Impaired Peripheral Conversion of Thyroxine to Triiodthyronine	R. R. Cavalieri, B. Rapoport	28:57–65
Diagnosis of Hyperparathyroidism	H. C. Bone III, W. H. Snyder, C. Y. C. Pak	28:111–17
The Role of Glucagon in the Endogenous Hyperglycemia of Diabetes Mellitus	R. H. Unger, L. Orci	28:119–30
Paget's Disease of Bone	W. G. Ryan	28:143–52
Classification of the Hyperlipidemias	R. J. Havel	28:195–209
Prostaglandins and Hypercalcemic States	H. W. Seyberth, L. G. Raisz, J. A. Oates	29:23–29
Clinical Hypophosphatemia	F. Fitzgerald	29:177–89
Osteomalacia and Disorders of Vitamin D Metabolism	J. F. Habener, J. E. Mahaffey	29:327–42
Humoral Syndromes Associated with Cancer	W. D. Odell, A. R. Wolfsen	29:379–406
Medical Treatment of Acromegaly	I. D. Goldfine	29:407–15
Control of the Onset of Puberty	R. M. Boyar	29:509–20
The Syndrome of Hyporeninemic Hypoaldosteronism	J. P. Knochel	30:145–53
Dietary Management of Diabetes Mellitus	J. W. Ensinck, E. L. Bierman	30:155–70
Big ACTH and Bronchogenic Carcinoma	R. S. Yalow	30:241–48
Prolactin and the Diagnosis of Pituitary Tumors	W. B. Malarkey	30:249–58
Diagnosis and Medical Management of Insulinomas	S. S. Fajans, J. C. Floyd Jr.	30:313–29
Diabetic Ketoacidosis: Reappraisal of Therapeutic Approach	A. E. Kitabchi, R. Young, H. Sacks, L. Morris	30:339–57
Hemoglobin A_{Ic} and Diabetes Mellitus	R. J. Koenig, A. Cerami	31:29–34
Efficacy and Safety of Oral Hypoglycemic Agents	H. S. Seltzer	31:261–72
Treatment of Hyperthyroidism	M. R. McClung, M. A. Greer	31:385–404
Bartter's Syndrome	J. R. Gill	31:405–19

GASTROINTESTINAL SYSTEM

Stress Ulceration in the Acutely Ill	J. J. Skillman, W. Silen	27:9–22
Infectious Particles in Hepatitis	A. P. Waterson	27:23–35
New Observations in Crohn's Disease	H. D. Janowitz, D. B. Sachar	27:269–85
Reye's Syndrome	G. M. Schiff	27:447–52
Viral Agents and Diarrheal Illness	H. L. DuPont, L. Portnoy, R. B. Conklin	28:167–77
Applications and Hazards of Total Intravenous Hyperalimentation	S. J. Dudrick, J. M. Long III	28:517–28
Digestion and Absorption of Protein	H. J. Freeman, Y. S. Kim	29:99–116
Gastrointestinal Hormones	J. E. McGuigan	29:307–18
Non-A, Non-B Hepatitis	S. M. Feinstone, R. H. Purcell	29:359–66
Cellular Mechanisms in Acid Secretion	A. H. Soll, M. I. Grossman	29:495–507
Angiographic Management of Gastrointestinal Bleeding	S. Kadir, C. A. Athanasoulis	30:41–59
The Medical Management of Peptic Ulcer	D. J. C. Shearman, D. Hetzel	30:61–79
Computed Tomography of the Hepatobiliary System	M. Korobkin, H. I. Goldberg	30:181–88
Amanita Mushroom Poisoning	D. H. Mitchel	31:51–57
The Medical Treatment of Gallstones	I. A. D. Bouchier	31:59–77
Intestinal Absorption and Malabsorption of Folates	C. H. Halsted	31:79–87
Flatulence	M. D. Levitt, J. H. Bond	31:127–37
Folate-Binding Proteins	N. Colman, V. Herbert	31:433–39

HEMATOLOGY

Hemochromatosis: Course and Treatment	S. Sherlock	27:143–49
Vitamin K and the Oral Anticoagulant Drugs	R. A. O'Reilly	27:245–61
Medical Management of Venous Thrombosis	S. Wessler	27:313–19
Factors in the Pathomechanism of Chronic Lymphocytic Leukemia	H. Theml, R. Love, H. Begemann	28:131–41
Paroxysmal Nocturnal Hemoglobinuria (PNH) as a Clonal Disorder	R. C. Hartmann, A. B. Arnold	28:187–94
Cold Agglutinins and Cold-Agglutinin Disease	M. M. Frank, J. P. Atkinson, J. Gadek	28:291–98
Treatment of Idiopathic Thrombocytopenic Purpura (ITP)	Y. S. Ahn, W. J. Harrington	28:299–309
Relationships Between Platelets and Coagulation Factors in Hemostasis	P. W. Majerus, J. P. Miletich	29:41–49
Erythropoietin and the Control of Red Cell Production	S. E. Graber, S. B. Krantz	29:51–66
Heparin, Antithrombin, and Abnormal Clotting	R. D. Rosenberg	29:367–78
Contractile Proteins in Cell Structure and Function	T. P. Stossel	29:427–57
Hairy Cell Leukemia	R. C. Braylan, J. S. Burke	30:17–24
Secretable Storage Pools in Platelets	H. Holmsen, H. J. Weiss	30:119–34
Disseminated Intravascular Coagulation: A Reappraisal	R. W. Colman, S. J. Robboy, J. D. Minna	30:359–74
Treatment of Plasma Cell Myeloma	D. E. Bergsagel	30:431–43
Prostaglandins and Platelets	J. F. Mustard, R. L. Kinlough-Rathbone, M. A. Packham	31:89–96
Treatment of Chronic Myeloid Leukemia	P. A. Bryckmans	31:159–66
Autoimmune Granulocytopenia	G. L. Logue, D. S. Shimm	31:191–200
Erythropoiesis	C. Peschle	31:303–14
The Endothelium and Thrombosis	R. T. Wall, L. A. Harker	31:361–71
Folate-Binding Proteins	N. Colman, V. Herbert	31:433–39
Controversies in Platelet Transfusion Therapy	S. J. Slichter	31:509–40

IMMUNOLOGY

The Association of Genes in the Major Histocompatibility Complex and Disease Susceptibility	T. Sasazuki, F. C. Grumet, H. O. McDevitt	28:425–52

Control of the Immune System by Inhibitor and Inducer T Lymphocytes	H. Cantor	30:269–77
Immunobiology of the Maternal-Fetal Relationship	R. E. Rocklin, J. L. Kitzmiller, M. D. Kaye	30:375–404

INFECTIOUS DISEASES

Biology of Endotoxin	R. J. Elin, S. M. Wolff	27:127–40
Significance of Serum Protein and Tissue Binding of Antimicrobial Agents	W. A. Craig, C. M. Kunin	27:287–300
Immunoprophylaxis of Gram-Negative Bacillary Infections	W. R. McCabe	27:335–41
Cytomegalovirus (CMV) in the Compromised Host(s)	R. F. Betts, J. B. Hanshaw	28:103–10
Pertussis in the Adult	C. C. Linnemann Jr., J. Nasenbeny	28:179–85
The Emergence of Group B Streptococci in Infections of the Newborn Infant	B. F. Anthony, D. M. Okada	28:355–69
The Current Status of BCG Immunization Against Tuberculosis	T. C. Eickhoff	28:411–23
E. Coli Meningitis: K_1 Antigen and Virulence	C. M. Wilfert	29:129–36
Otitis Media: The Spectrum of Middle Ear Inflammation	G. S. Geibink, P. G. Quie	29:285–306
Rabies Vaccination	S. A. Plotkin, T. Wiktor	29:583–91
Immunization Against Typhoid Fever	J. W. Warren, R. B. Hornick	30:457–72
Legionnaires' Disease	R. D. Meyer, S. M. Finegold	31:219–32
Pulmonary Function During and After Common Respiratory Diseases	W. J. Hall, R. G. Douglas, Jr.	31:233–38
Giardiasis	J. W. Smith, M. S. Wolfe	31:373–83
Active Immunizations for Adults	G. M. Schiff	31:441–51
Infant Botulism	S. S. Arnon	31:541–60

KIDNEY AND URINARY SYSTEM

Late Sequelae of Poststreptococcal Glomerulonephritis	D. S. Baldwin, R. G. Schacht	27:49–55
Glomerular Injury in Patients with Neoplasia	B. S. Kaplan, J. Klassen, M. H. Gault	27:117–25
The Effects of Calcium and Phosphorus in Hemodialysis	R. S. Goldsmith	27:181–90
Interrelations of Renin, Angiotensin II, and Sodium in Hypertension and Renal Failure	H. Gavras, J. A. Oliver, P. J. Cannon	27:485–521
IgA-Associated Glomerulonephritis	J. J. McPhaul Jr.	28:37–42
Glomerular Filtration of Fluid and Macromolecules: The Renal Response to Injury	D. A. Maddox, B. M. Brenner	28:91–102
Renal Function During and Following Obstruction	D. R. Wilson	28:329–39
Urate Excretion	G. M. Fanelli Jr.	28:349–54
Recurrent Glomerulonephritis After Renal Transplantation	J. Hamburger, J. Crosnier, L.-H. Noël	29:67–72
Dialysis Encephalopathy Syndrome	A. C. Alfrey	29:93–98
Immunological Aspects of Renal Tubular and Interstitial Diseases	R. T. McCluskey, R. B. Colvin	29:191–203
Chronic Bacteremia and Nephritis	Y. Kim, A. F. Michael	29:319–25
Dialysis Versus Transplantation in the Treatment of End-Stage Renal Disease	J. P. Merrill	29:343–58
Renal Tubular Acidosis	F. J. Gennari, J. J. Cohen	29:521–41
Low Molecular Weight Proteinuria	C. L. Hall, J. Hardwicke	30:199–211
The Ultrastructure of Glomerular Filtration	M. J. Karnovsky	30:213–24
Localization and Antibiotic Management of Urinary Tract Infection	L. S. T. Fang, N. E. Tolkoff-Rubin, R. H. Rubin	30:225–39
Prostatitis	E. M. Meares, Jr.	30:279–88
Patterns of Circulating Complement in Renal Diseases	E. J. Lewis	30:445–55

Salt Balance and Long-Term Blood Pressure Control	A. C. Guyton, T. G. Coleman, D. B. Young, T. E. Lohmeier, J. W. DeClue	31:15–27
Control of Acid Excretion by the Kidney	R. L. Tannen	31:35–49
Plasma Exchange in Glomerulonephritis and Related Vasculitides	C. M. Lockwood, D. K. Peters	31:167–79
Idiopathic Membranoproliferarive Glomerulonephritis	Y. Kim, A. F. Michael	31:273–88
Vasopressin Function in the Syndrome of Inappropriate Antidiuresis	R. Zerbe, L. Stropes, G. Robertson	31:315–27
Pericarditis and Renal Failure	R. Renfrew, T. J. Buselmeier, C. M. Kjellstrand	31:345–60
The Natural History and Response to Therapy of Lupus Nephritis	E. M. Ginzler, A. J. Bollet, E. A. Friedman	31:463–87

MISCELLANEOUS

Some Implications of Population Control	A. C. Barnes	27:1–7
Potential Microwave Injuries in Clinical Medicine	D. I. McRee	27:109–15
Implications of Mass Automated Instruments on Medical Practice	R. R. Grams	27:199–206
Medical Aspects of Riot Control (Harassing) Agents	J. P. Sanford	27:421–29
Confidentiality in Medical Practice	M. Grossman	28:43–55
Informed Consent	G. J. Annas	29:9–14
The XYY Genotype	S. Walzer, P. S. Gerald, S. A. Shah	29:563–70
Recent Membrane Research and Its Implications for Clinical Medicine	V. T. Marchesi	29:593–603
The Efficacy of Biofeedback Therapy	M. T. Orne	30:489–503
Snake Venom Poisoning in the United States	F. E. Russell	31:247–59

NEOPLASIA

Tumor Antigens and Tumor-Host Relationships	R. W. Baldwin, M. R. Price	27:151–63
The Functions of the Macrophage in Malignant Disease	P. Alexander	27:207–24
Multiple Endocrine Neoplasia Type II	S. A. Wells Jr., D. A. Ontjes	27:263–68
The Role of Oncogenic Viruses in Neoplasia	P. A. Klein, R. T. Smith	28:311–27
Structure and Function of Alpha-Fetoprotein	T. B. Tomasi Jr.	28:453–65
Surgical Adjuvant Chemotherapy	M. Perloff, J. F. Holland	28:475–88
BCG as Adjuvant Immunotherapy for Neoplasia	E. M. Hersh, J. U. Gutterman, G. M. Mavligit	28:489–515
The Non-Hodgkin's Lymphomas: Current Concepts and Management	C. S. Portlock, E. Glatstein	29:81–91
Chemotherapy in Prevention of the Recurrence of Resectable Cancer	G. Bonadonna, G. Tancini, A. Rossi, M. Gasparini	29:149–75
Immunotherapy for Malignant Disease	J. E. Goodnight Jr., D. L. Morton	29:231–83
Biological and Enzymatic Events in Chemical Carcinogenesis	H. C. Pitot	30:25–39
Clonal Origin of Human Tumors	P. J. Fialkow	30:135–43
Growth Considerations in the Radiation Therapy of Children with Cancer	P. S. Rittman, G. J. D'Angio	30:405–15

NERVOUS SYSTEM

Temporal Lobe Epilepsy: Neurological and Behavioral Aspects	I. Sherwin	27:37–47
Computed Tomography in Neurologic Diagnosis	A. G. Osborn	30:189–98
Prenatal Sex Hormones and the Developing Brain: Effects on Psychosexual Differentiation and Cognitive Function	A. A. Ehrhardt, H. F. L. Meyer-Bahlburg	30:417–30

PSYCHIATRY		
Reassessment of Homosexuality and Transsexualism	R. C. Friedman, R. Green, R. L. Spitzer	27:57–62
The Relation of Psychological Stress to Onset of Medical Illness	M. W. Hurst, C. D. Jenkins, R. M. Rose	27:301–12
Behavioral and Biochemical Interrelations in Alcoholism	J. H. Mendelson, N. K. Mello	27:321–33
Psychological Assistance for the Dying Patient and His Family	T. P. Hackett	27:371–78
Neuroendocrine Dysfunction in Depressive Illness	E. J. Sachar	27:389–96
The Sleep Apnea Syndromes	C. Guilleminault, A. Tilkian, W. C. Dement	27:465–84
New Concepts in the Diagnosis of Psychiatric Disorders	E. Robins	28:67–73
Anorexia Nervosa: Recent Investigations	K. A. Halmi	29:137–48
Specificity of Depressive Diseases	D. R. Sweeney, J. W. Maas	29:219–29
Genetic Transmission of Schizophrenia	D. K. Kinney, S. Matthysse	29:459–73
Management of Insomnia	C. R. Soldatos, A. Kales, J. D. Kales	30:301–12
REPRODUCTIVE SYSTEM		
Cardiovascular Complications of Oral Contraceptives	N. M. Kaplan	29:31–40
The Prenatal Diagnosis of Genetic Disorders	C. J. Epstein, M. S. Golbus	29:117–28
Endocrinology of the Placenta	R. B. Everett, P. C. MacDonald	30:473–88
Effectiveness and Risks of Contraception	W. Droegemueller, R. Bressler	31:329–43
RESPIRATORY SYSTEM		
The Adult Respiratory Distress Syndrome	P. C. Hopewell, J. F. Murray	27:343–56
Hypersensitivity Pneumonitis: Current Concepts of Etiology and Pathogenesis	M. Lopez, J. Salvaggio	27:453–63
Diagnosis and Treatment of Fat Embolism	J. A. Moylan, M. A. Evenson	28:85–90
Diagnosis of Pulmonary Embolism	G. V. R. K. Sharma, D. E. Tow, A. F. Parisi, A. A. Sasahara	28:159–66
Biology of Drowning	J. H. Modell	29:1–8
Clinical Problems of Ventilatory Control	G. B. Irsigler, J. W. Severinghaus	31:109–26
Pulmonary Function During and After Common Respiratory Diseases	W. J. Hall, R. G. Douglas Jr.	31:233–38
Primary Pulmonary Hypertension	A. P. Fishman, G. G. Pietra	31:421–31
RHEUMATOLOGY		
Polymyalgia Rheumatica and Giant Cell Arteritis	R. E. Ettlinger, G. C. Hunder, L. E. Ward	29:15–22
SURGICAL TOPICS		
Parietal Cell Vagotomy Without Drainage for the Treatment of Duodenal Ulcer	P. H. Jordan Jr.	27:191–97
Metabolic Complications of Jejunoileal Bypass Operations for Morbid Obesity	H. W. Scott Jr., R. H. Dean, H. J. Shull, F. W. Gluck	27:397–405
Corneal Transplantation	S. L. Forstot, H. E. Kaufman	28:21–35
Pulmonary Complications of General Surgery	A. K. Pierce, J. Robertson	28:211–21
Critique of Coronary Artery Bypass Surgery	C. B. Mullins, K. Lipscomb	28:271–89
Vitrectomy Surgery for Diabetic Retinopathy	G. H. Bresnick, F. L. Myers	30:331–38

R FORM ANNUAL REVIEWS INC.

st on the order blank on the reverse side the volumes you wish to order and whether you wish a standing order
atest volume sent to you automatically upon publication each year). Volumes not yet published will be shipped in
nd year indicated. Prices subject to change without notice. Out of print volumes subject to special order.

W TITLES FOR 1980

UAL REVIEW OF PUBLIC HEALTH ISSN 0163-7525
Vol. 1 (avail. May 1980): $17.00 (USA), $17.50 (elsewhere) per copy

NUAL REVIEWS REPRINTS: IMMUNOLOGY, 1977–1979 ISBN 0-8243-2502-8
ollection of articles reprinted from recent *Annual Review* series
Avail. Mar. 1980 Soft cover: $12.00 (USA), $12.50 (elsewhere) per copy

CIAL PUBLICATIONS

UAL REVIEWS REPRINTS: CELL MEMBRANES, 1975–1977 ISBN 0-8243-2501-X
ection of articles reprinted from recent *Annual Review* series
ublished 1978 Soft cover: $12.00 (USA), $12.50 (elsewhere) per copy

XCITEMENT AND FASCINATION OF SCIENCE, VOLUME 1 ISBN 0-8243-1602-9
lection of autobiographical and philosophical articles by leading scientists
Published 1965 Clothbound: $6.50 (USA), $7.00 (elsewhere) per copy

EXCITEMENT AND FASCINATION OF SCIENCE, VOLUME 2: Reflections by Eminent Scientists
Published 1978 Hard cover: $12.00 (USA), $12.50 (elsewhere) per copy ISBN 0-8243-2601-6
Soft cover: $10.00 (USA), $10.50 (elsewhere) per copy ISBN 0-8243-2602-4

HISTORY OF ENTOMOLOGY ISBN 0-8243-2101-7
ial supplement to the *Annual Review of Entomology* series
Published 1973 Clothbound: $10.00 (USA), $10.50 (elsewhere) per copy

NUAL REVIEW SERIES

ual Review of ANTHROPOLOGY ISSN 0084-6570
Vols. 1–8 (1972–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 9 (avail. Oct. 1980): $20.00 (USA), $21.00 (elsewhere) per copy

ual Review of ASTRONOMY AND ASTROPHYSICS ISSN 0066-4146
Vols. 1–17 (1963–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 18 (avail. Sept. 1980): $20.00 (USA), $21.00 (elsewhere) per copy

ual Review of BIOCHEMISTRY ISSN 0066-4154
Vols. 28–48 (1959–79): $18.00 (USA), $18.50 (elsewhere) per copy
Vol. 49 (avail. July 1980): $21.00 (USA), $22.00 (elsewhere) per copy

ual Review of BIOPHYSICS AND BIOENGINEERING* ISSN 0084-6589
Vols. 1–8 (1972–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 9 (avail. June 1980): $17.00 (USA), $17.50 (elsewhere) per copy

nual Review of EARTH AND PLANETARY SCIENCES* ISSN 0084-6597
Vols. 1–7 (1973–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 8 (avail. May 1980): $17.00 (USA), $17.50 (elsewhere) per copy

nnual Review of ECOLOGY AND SYSTEMATICS ISSN 0066-4162
Vols. 1–10 (1970–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 11 (avail. Nov. 1980): $20.00 (USA), $21.00 (elsewhere) per copy

nnual Review of ENERGY ISSN 0362-1626
Vols. 1–4 (1976–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 5 (avail. Oct. 1980): $20.00 (USA), $21.00 (elsewhere) per copy

Annual Review of ENTOMOLOGY* ISSN 0066-4170
Vols. 7–24 (1962–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 25 (avail. Jan. 1980): $17.00 (USA), $17.50 (elsewhere) per copy

Annual Review of FLUID MECHANICS* ISSN 0066-4189
Vols. 1–11 (1969–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 12 (avail. Jan. 1980): $17.00 (USA), $17.50 (elsewhere) per copy

Annual Review of GENETICS ISSN 0066-4197
Vols. 1–13 (1967–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 14 (avail. Dec. 1980): $20.00 (USA), $21.00 (elsewhere) per copy

Annual Review of MATERIALS SCIENCE ISSN 0084-6600
Vol. 1–9 (1971–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 10 (avail. Aug. 1980): $20.00 (USA), $21.00 (elsewhere) per copy

(continued on reverse)

ice will be increased to $20.00 (USA), $21.00 (elsewhere) per copy effective with the 1981 volume.

Annual Review of MEDICINE: Selected Topics in the Clinical Sciences* ISSN
Vols. 1–3, 5–15, 17–30 (1950–52, 1954–64, 1966–79): $17.00 (USA), $17.50 (elsewhere)
Vol. 31 (avail. Apr. 1980): $17.00 (USA), $17.50 (elsewhere) per copy

Annual Review of MICROBIOLOGY ISSN 0
Vols. 15–33 (1961–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 34 (avail. Oct. 1980): $20.00 (USA), $21.00 (elsewhere) per copy

Annual Review of NEUROSCIENCE* ISSN 0
Vols. 1–2 (1978–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 3 (avail. Mar. 1980): $17.00 (USA), $17.50 (elsewhere) per copy

Annual Review of NUCLEAR AND PARTICLE SCIENCE ISSN 0
Vols. 10–29 (1960–79): $19.50 (USA), $20.00 (elsewhere) per copy
Vol. 30 (avail. Dec. 1980): $22.50 (USA), $23.50 (elsewhere) per copy

Annual Review of PHARMACOLOGY AND TOXICOLOGY* ISSN
Vols. 1–3, 5–19 (1961–63, 1965–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 20 (avail. Apr. 1980): $17.00 (USA), $17.50 (elsewhere) per copy

Annual Review of PHYSICAL CHEMISTRY ISSN
Vols. 10–21, 23–30 (1959–70, 1972–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 31 (avail. Nov. 1980): $20.00 (USA), $21.00 (elsewhere) per copy

Annual Review of PHYSIOLOGY* ISSN
Vols. 18–41 (1956–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 42 (avail. Mar. 1980): $17.00 (USA), $17.50 (elsewhere) per copy

Annual Review of PHYTOPATHOLOGY ISSN
Vols. 1–17 (1963–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 18 (avail. Sept. 1980): $20.00 (USA), $21.00 (elsewhere) per copy

Annual Review of PLANT PHYSIOLOGY* ISSN
Vols. 10–30 (1959–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 31 (avail. June 1980): $17.00 (USA), $17.50 (elsewhere) per copy

Annual Review of PSYCHOLOGY* ISSN
Vols. 4, 5, 8, 10–30 (1953, 1954, 1957, 1959–79): $17.00 (USA), $17.50 (elsewhere) per
Vol. 31 (avail. Feb. 1980): $17.00 (USA), $17.50 (elsewhere) per copy

Annual Review of SOCIOLOGY ISSN
Vols. 1–5 (1975–79): $17.00 (USA), $17.50 (elsewhere) per copy
Vol. 6 (avail. Aug. 1980): $20.00 (USA), $21.00 (elsewhere) per copy

*Price will be increased to $20.00 (USA), $21.00 (elsewhere) per copy effective with the 1981 volume.

To ANNUAL REVIEWS INC., 4139 El Camino Way, Palo Alto, CA 94306 USA
(Tel. 415-493-4400)

Please enter my order for the following publications:
(Standing orders: indicate which volume you wish order to begin with)

______________________________, Vol(s). ____ Standing o
______________________________, Vol(s). ____ Standing order ____
______________________________, Vol(s). ____ Standing o
______________________________, Vol(s). ____ Standing ord

Amount of remittance enclosed $________ California residents please add applicable
Please bill me ☐ Prices subject to change withour

SHIP TO (include institutional purchase order if billing address is different)

Name ______________________________

Address ______________________________

____

______________________________ Zip Code ____________

Signed ______________________________ Date ____________

☐ Please add my name to your mailing list to receive a free copy of the current Prospectus e
☐ Send free brochure listing contents of recent back volumes for *Annual Review(s)* of

____